Chapter 11. Fire Alarms and Democratic Accountability

Charles M. Cameron and Sanford C. Gordon
1 Introduction

Accountability is a situation that prevails in any relationship between a principal and agent in which the latter takes some action for which she may be “held to account” – that is, made to answer for those actions under a system of rewards and sanctions administered by the former. In a democratic setting, the electorate are the principals and public officials the agents; rewards and sanctions the loss or retention of the benefits, privileges, and powers of office-holding. Absent electoral accountability, citizens must fall back on the benevolence of their rulers or just plain luck. History shows these are weak reeds indeed. Thus, the accountability (or unaccountability) of elected officials to voters is a key component of any comprehensive theory of democratic governance.

Electoral accountability faces huge obstacles in practice. One of the most pernicious is the problem of asymmetric information: while knowing what politicians are up to is surely a critical ingredient of holding incumbent officeholders to account, for most voters, becoming and remaining informed about whether elected officials are actually meeting their obligations can be tedious, time-consuming and difficult. How many citizens have the leisure to research the voting record of their member of Congress, for instance? And even then, outside of a few specific areas, how many have the knowledge to assess whether a given legislative enactment actually improved their welfare, all things considered? A lack of transparency about actions and the obscurity between means and ends can render nominal accountability completely moot in practice.\(^1\)

In the face of prohibitive information costs, a simple intuition is the following: any mechanism that lowers informational costs for voters ought to enhance electoral accountability. And, greater electoral accountability should make voters better off. Unfortunately, recent advances in the analysis of electoral accountability shows that this simple intuition can be more problematic than it seems at first blush. For example, when a politician’s actions are readily observable but their consequences delayed or obscure – the “wrong kind of transparency” – politicians may have an incentive to “pander” – that is, to take popular actions today even when they know full well the consequences may

\(^1\)To make matters worse, informed voters supply a public good – knowledgeable oversight of politicians – to uninformed citizens. If gathering information is costly for citizens, then members of the electorate face a collective action problem in which the uninformed can free-ride on the efforts of the informed. The resulting incentives may result in little information gathering and poor oversight of politicians [Hardin 2006]. But if information about politician behavior is effectively costless, then informational free-riding in the electorate becomes less concerning.
prove ineffectual or even harmful to voters. ([Canes-Wrone et al 2001], [Maskin and Tirole 2004], see also [Prat 2005], [Fox and van Weelden 2012]). Under such conditions, a degree of opacity concerning a politician’s actions, if not the consequences of her choices, may soften the incentive to pander and actually improve the lot of voters. The lesson is that one must be very careful about overly simple intuitions when it comes to informational transparency and electoral accountability.

Nonetheless, the sense remains strong that if voters had access to mechanisms that lower their information costs – at least for the “right” kind of information – then democratic accountability might be enhanced, to the benefit of voters. What might those mechanisms be?

1.1 Fire Alarms

One important cost-lowering mechanism is fire alarms. In the context of accountability between a worker and a boss, a “fire alarm” connotes a readily perceived, reliable notice about bad worker performance that – critically – comes free or nearly free to the boss herself. In an electoral setting, a fire alarm consists of such a notice to voters about the performance of an incumbent. Might fire-alarms snatch electoral accountability from the jaws of information costs? That is our subject in this chapter.

Among political scientists, the idea of fire alarms gained currency from a celebrated analysis of congressional oversight of the bureaucracy ([McCubbins and Schwartz 1984]). In this setting, Congress (the boss or principal) has a hard time perceiving or evaluating the actions of bureaucrats (the workers or agent). McCubbins and Schwartz note that administrative procedures provide opportunities for private interests to relay information to Congress about bureaucratic noncompliance with legislative preferences. Relying on these actors – whom we will refer to generically as sentinels – is more efficient for Congress, they argue, than undertaking costly, active “police patrol” oversight. Thus, passive reliance on fire alarms need not imply congressional abdication of its oversight responsibility. Moreover, anticipation of the fire alarm may deter bureaucratic deviation from congressional desires. It seems plausible, then, that accountability can work well despite information costs, so long as some actor is willing to bear the informational costs for the boss.

Yet all may not be copacetic in the world of bureaucratic oversight. What happens when the sentinel is biased? Prendergast’s ([Prendergast 2003]) game-theoretic analysis of service-providing bureaucracies analyzes a strategic situation closely related to the one discussed by McCubbins and
Schwartz, but in which the sentinel is not neutral in the signals it sends to a bureaucrat’s political bosses. Much as students will only complain to their professors about an unexpectedly low grade conferred by a teaching assistant (but not an unexpectedly high one), so too will the bureaucrat’s client only complain when a service or benefit is denied (but not when one is granted – justifiably or not). Prendergast explores how this asymmetry affects the behavior of bureaucrats, and how their political principals may respond to neutralize the distortion. The overall lesson is clear: one needs to pay attention to sentinel bias.

Political scientists have explored the logic of fire-alarm oversight in other settings. A notable one is Supreme Court oversight of decisions in the U.S. Courts of Appeals. In this setting, the lower court (the “worker”) is a three-judge panel deciding a case. The panel may do so in conformance with the preferred doctrine of the Supreme Court (the “boss”), but it may also deviate from the high court’s preferred doctrine. For the high court, detecting such a deviation is difficult because, absent review, it is not privy to all the information available to the lower court. In an influential analysis, Cross and Tiller noted the following: if one of the judges on the panel is aligned with the higher court, she may announce a deviation via a dissenting vote (Cross and Tiller 1998). The dissenting vote from its ally is thus a fire alarm for the higher court, alerting it to the deviation and allowing it to review and correct the lower court’s action. And, Cross and Tiller note, the possibility of fire-alarm dissents may in turn deter doctrinal deviations. Thus, fire-alarm oversight in the judicial hierarchy would seem to afford an easy path to judicial consistency and compliance with the rule of law.

Yet a caveat is once again in order. In a recent analysis, a group of scholars has returned to fire-alarm oversight in the judiciary, applying careful game theoretic reasoning to Cross and Tiller’s informal arguments (Beim, Hirsch, and Kastellec 2014). Their path-breaking analysis confirms the intuitions of Cross and Tiller in some ways but also uncovers a potential issue: a “boy who cried wolf” problem. If the interests of the sentinel judge on the panel are too extreme, she may sound alarms too frequently, trying to induce the high court to review marginal deviations that it would prefer to let slide given the effort of reversing them. Crying wolf too often can lead the high court to ignore the dissenter’s fire-alarm, which obviates the impact of fire-alarm oversight. Although this problem is different from those found in Prendergast’s analysis of bureaucrats and their clients, the lesson is similar: one needs to be very attentive to the interests of the sentinel.
The logic of fire alarms oversight would seem to extend naturally to an electoral context. Here, the seminal analysis was offered by Arnold ([Arnold 1993]). In an expansion of his argument in *The Logic of Congressional Action* [Arnold 1993], Arnold noted that fire alarms arise in the principal-agent relationship between voters and elected officials. He argued that political activists may have, and electoral challengers certainly do have, strong incentives to sound an alarm in the event of legislative malfeasance or error by a member of Congress. It is the threat of fire alarms, in this view, that might motivate legislator compliance with the preferences of inattentive citizens, just as the threat of fire alarms may induce bureaucrats to abide by the preferences of an inattentive Congress and lower court judges to attend to the doctrinal preferences of an over-burdened higher court.

It worth reviewing Arnold’s argument at length:

The system contains activists who have incentives to monitor what legislators are doing in office and to inform citizens when legislators fail in their duties. Challengers to incumbent legislators have perhaps the strongest incentives for monitoring legislators’ behavior and mobilizing voters. Few challengers fail to sift through incumbent voting records in search of issues that can be used against incumbent legislators. In addition, groups that bear major costs under a particular governmental policy may help publicize what incumbent legislators have done to contribute to their plight. Whereas challengers seek to replace incumbents, these groups may seek to persuade incumbents to avoid electoral repercussions by altering their positions and working for the groups’ benefit ... Uncertainty abounds in a system like this. Legislators cannot possibly know for sure what policy effect will follow from specific governmental actions, how challengers or interest group leaders might use governmental actions or inactions to stir up citizens, or whether citizens might blame or absolve legislators for their connections to specific actions. What is certain is that legislators will do their best to anticipate citizens’ preferences, to avoid the most dangerous mine fields, and to chart as safe a course as possible through the treacherous territory before them ... When legislators adjust their voting decisions to avoid generating preferences among inattentive citizens, is it fair to suggest that legislators are controlled by those inattentive citizens? It is indeed.
The critical point Arnold is making is that activists and challengers serve as sentinels, and that in this role their presence may enhance electoral accountability. As in the bureaucratic and judicial examples, the intuition is appealing. But the careful and subtle analyses of Prendergast in the bureaucratic setting and of Beim, Hirsch, and Kastellec in the judicial setting sound a warning: careful attention to sentinel’s incentives is surely essential in mapping out the promise and perils of fire alarms for democratic accountability.

1.2 This Chapter

The aim of this chapter is to open the black-box of electoral fire alarms. We do not claim to offer more than a preliminary sounding of some rather deep waters. Nonetheless we assay some provocative findings. In our analysis, the incentives of activists and those of challengers are quite different. As a result, fire-alarms from activists display many of the problems identified by Prendergast’s analysis of bureaucratic fire-alarms. In contrast, fire-alarms from challengers have the potential to be extremely informative to voters, especially in tandem with information from incumbents. But there is an enormous caveat: the fire-alarm information must be nearly costlessly verifiable for voters. Absent “hard” information, challenger fire-alarms are vulnerable to attacks as “fake,” “phony,” or “fraud” – and often will be. Alarms believed to be fake are valueless to voters. This caveat highlights the extreme importance of visible and believable sources of fact-checking, like trusted media outlets and investigative journalism. This caveat might have seemed of mostly theoretic interest when Arnold first raised the idea of electoral fire alarms in the early 1990s. But in today’s American politics, its appears disturbingly relevant.

The chapter is laid out as follows. In the next section, we provide a simple review of modern electoral accountability theory as it bears on fire-alarm oversight of politicians. We are very selective as this literature has become enormously bulky and forbiddingly complex. Yet in the area of electoral fire-alarms it is sparse. We then turn to a very simple model of electoral fire-alarms. Our presentation is non-technical, emphasizing intuition, though more details are available elsewhere. The fourth section of the chapter offers some mini-cases of congressional elections and fire-alarm

---

2And, we are not the first. We review [Ashworth and Shotts 2011], the closest work, below. [Gordon and Huber 2007] contains an informal theoretical discussion that discusses the role of the challenger as fire alarm, and the incentives this creates for the incumbent.
accountability. These illustrates some of the themes flagged by the analysis. We conclude in the final section.

2 Sentinels and the Theory of Political Agency

2.1 The Voter and the Official

Our analysis of sentinels in the next section draws heavily on recent theoretical advances in the theory of political agency (TPA; see Ashworth 2012 and Gailmard 2014 for a review). Although our presentation is not especially technical, it may be tough going for those lacking familiarity with this branch of political economy. So, before tackling the analysis, we'll review a few of the basic ideas and building blocks in TPA, consider how sentinels fit into the standard account, and glance at related studies within the TPA tradition.

For students approaching the study of accountability from the perspective of the mainstream political science research on the U.S., the first thing to understand is that TPA owes very little to classics in that literature from such scholars as David Mayhew, Richard Fenno, or R. Douglas Arnold. Likewise, if your understanding of the relationship between citizen and elected officials draws critical distinctions between prospective and retrospective voting; delegate and trustee models of representation; or substantive versus descriptive representation; you will find little familiar and much that is missing, at least at first glance.

One critical reason for the disjuncture is that those studies tend to focus on legislatures and legislators. A critical feature of legislatures, of course, is that they are bodies consisting of multiple agents. And thus, much of the literature on accountability and representation has focused on important features of this multiplicity: inter alia, the internal organization of the legislature, parties and party leadership, seniority, coalition building, and team production of legislation. By contrast, scholars working in the TPA tradition – borrowing heavily from contract theory – have tended to abstract away from the multiplicity of legislators, focusing instead on the relationship between a single elected official (the incumbent) and an electorate that must choose periodically whether to retain her. Many models go even further, abstracting away from an electorate of many, heterogeneous voters in favor of a single, representative voter (for example, the median). As with

---

3 For readers with some technical facility, Besley 2006 and Fearon 1999 are clear and thoughtful introductions to TPA. Duggan and Martinelli 2017 is rewarding for more advanced readers.

4 For canonical research in these areas, see, e.g., Pitkin 1967, Key 1966, and Fiorina 1981.
all modeling enterprises, the point of this stark abstraction is not to deny the importance of the institutional and behavioral richness to be found in legislative settings. Rather, it is to hold fixed one set of relationships in the political environment in order to focus and gain insight on another. This stripped down approach offers an extremely clean way to consider some genuinely deep issues about elections and accountability, and accordingly, it is the approach we adopt in what follows. That being said, one could argue that the elected official envisioned by many models in the TPA tradition is more akin to an executive like a president, governor, or mayor, who can take some form of unilateral action that may be observed by the voters.

In the starkest models coming out of the TPA tradition, there are just two “players” – the incumbent official and the (representative) voter. A third player, the challenger, is often treated as a passive alternative – perhaps reflecting the intellectual origins of TPA approaches in contract theory, where replacing an agent entails a new draw from the labor market. Below, we will consider situations in which challengers are active players in their own right. An archetypal TPA game consists of two periods: in the first, an incumbent takes some action, which has some consequences for the voter; then, the voter may observe something – perhaps the action, perhaps the consequences (which one, as we will see, matters a lot) – and decides whether to retain the incumbent or replace her with a challenger. In the second period, the incumbent (or her replacement) again takes an action. This simple dynamic setting allows events and incentives to unfold in a natural way, and critically, introduces a motivation for voters to choose candidates they anticipate will act on their behalf in the second period. The elected official cares about holding office, either because she receives direct benefits from holding office, or because she has policy preferences that diverge from those of her opponent, such that losing would be costly.

This archetypal framework contains the essential aspects of an accountability relationship between official and voter: the official, in her capacity as the agent of the voter, takes some action for which she may be “held to account” under a system of punishment and reward administered by the citizens through their vote choices.

Given the sequence outlined above, the election dividing the two terms means that the voters

---

5Often, what will happen in the second period is so immediate that for convenience, the analyst will simply “roll up” the ensuing payoffs into first period payoffs and dispense with an explicit second period. We take this approach below, but the model is conceptually equivalent to one with two periods. One can, of course, extend beyond two periods, up to an including an infinite number.
casts ballots based on what they observe in the first period, with an eye toward what will happen (or is likely to happen) in the second. A voter in 2020 might, for example, as herself, “Given President Trump’s track record as I understand it, would he or his Democratic alternative likely prove a better president over the next four years?” and condition her vote on the answer to that question.

Hence, a critical insight of this literature is that the distinction between retrospective and prospective voting is really a distinction without a difference: voters form prospective forecasts of what will happen based on retrospective evaluations of what has happened. And what has happened is sunk cost, so voting against an incumbent to punish her for misdeeds in office is not sequentially rational for a voter who thinks the alternative is worse. To be sure, the sunk cost fallacy is a real thing. And so, in the real world, there are surely voters that cast their votes out of pure emotion – for example, out of vengeance against an incumbent whom they feel “crossed” them, even if it means cutting off their nose to spite their face. The TPA approach sets such voters aside, not because its practitioners believe they don’t exist, but rather with a specific analytical objective in mind: to isolate the incentives that may materialize for elected officials even in the absence of any concern about vengeful voters out to inflict punishment for past misdeeds.

Early TPA models (e.g., Ferejohn 1986) viewed politicians as basically interchangeable, and tending toward sloth or corruption. On this account, the election is a stick, and the fear of its stinging blows may coerce decent performance from the lazy, malfeasant shirkers we know all politicians to be. Critically, because politicians are interchangeable, the voter in this family of “pure moral hazard” models doesn’t really care about who occupies the office. Given this indifference, she can credibly commit to a schedule of electoral rewards that induces effort on the part of the incumbent: for example, “I will vote to reelect you if and only if your performance surpasses some threshold of performance.”

The implausibility of the incumbent homogeneity assumption that drives accountability in pure moral hazard models (see Fearon 1999 for an especially trenchant critique) has caused the vast

---

6 Note that if voters are driven by other emotion-based motivations, such as unreflective tribalism or intoxicated nationalist frenzy, this will serve to weaken electoral incentives.

7 More precisely, pure moral hazard models typically have many equilibria. The equilibrium selection criterion generally employed is the one that maximizes voter welfare, i.e. one that sanctions bad performance or behavior and rewards good performance. If these carrots and sticks are sufficient to induce good behavior, this accountability equilibrium hangs together.
majority of more recent research in the TPA tradition to take the heterogeneity of politicians as a
starting point. In this view, some politicians are “good types” from the viewpoint of the voters,
while others are “bad types” (where the meaning of good and bad may vary, depending on context,
as we explain momentarily). On this account, the election serves as an institutional mechanism
for selecting good types, which (hopefully) differentially retains good incumbents and weeds out
bad ones. Here, the politician’s actions (if observed) and the consequences of those actions (if
observed) may provide useful information about the politician’s type. At the outset, only the
politician knows his type for sure so the critical issue here is hidden information (the incumbent’s
type). Thus, the models seek to capture the voter’s “adverse selection” problem (the non-intuitive
language originates from early analyses of insurance). An example of a pure adverse selection model
is [Besley and Prat 2006], discussed below. One may combine both approaches, so the incumbent
may take a hidden action, with his preferred action depending on his nature.

In models with good and bad types of politicians, a key question is, what makes for a good
politician-employee and what for a bad one? To answer this question, it will be helpful to depart
momentarily from the principal-agent relationship between the voter and elected official, and con-
sider a more mundane relationship: one between you and your gardener. What makes for a good
gardener, whom you would like to retain, and what makes for a bad one you would wish replace?
An obviously desirable attribute for a gardener is skill: for example, a good gardener knows that
if the garden is dry it needs to be watered while if the garden is too wet it needs to dry out.
It is not implausible that voters would like similar skill in a president, governor, or mayor. In a
legislative context, expertise may come in the form of specific knowledge regarding the affect of
national policies in the incumbent’s district, or skill in navigating the legislative process to achieve
tangible benefits for one’s constituents.

A second desirable attribute for our gardener is congruence of preferences. For example, if you
like flowers and are allergic to peanuts, you want your gardener to grow the former and eschew the
latter. If he insists on growing deadly allergens, perhaps because he likes to eat them when you are
not around, then you should dismiss him and get someone who will happily tend gladiolas. Again, it
is highly plausible that voters want similar policy preferences in their president, governor, or mayor
since a congruent type will be inclined to ”do the right thing” from the voters’ perspective even
absent monitoring and sanctioning. And it is just as plausible that voters would want preference
congruence in a legislator, who will often represent them on votes they can neither observe (another hidden action) nor understand (more hidden information).

Many political scientists are quite familiar with policy preferences as modeled in the spatial theory of voting ([Enelow and Hinich 1984]). So, they will find it natural to think of a politician-employee having an "ideal point" on a line (perhaps as measured using metrics like NOMINATE scores), and the voter having a similar ideal point. On this account, choosing a politician-employee with similar policy preferences means choosing one whose ideal point is close to yours. Examples of PTA models with preferences like this include [Morelli and Van Weelden 2013], [Wolton 2019], and [Snyder and Ting 2003]; the preferences in [Maskin and Tirole 2004] can be interpreted this way as well.

What about skill, however? Here, PTA models often employ a convenient device to explore its benefits: state-contingent preferences with private signals. In this approach, there is a “state of the world” (e.g., the garden is dry or the garden is wet). A skilled employee can observe the state of the world and take the correct action that “matches” the state: e.g., water a dry garden, dry out a wet one. But an unskilled agent may be less certain about the state of the world and can only guess what to do. We formalize this by having the agent receive a private signal about the true state of the world; skilled agents receive accurate signals while unskilled ones receive noisy ones (or perhaps nothing at all). In Section 3, we will adopt the technology of state-contingent preferences and type-dependent private signals. For simplicity we ignore preference congruence, leaving that as a topic for future research.

A question of critical importance is what the voter can observe in the first term of office and what inferences she can draw on the basis of what she observed. An important distinction is between actions and consequences. In the gardening analogy, this is the distinction between seeing the gardener water (or not water) the plants and seeing whether the garden is thriving or blighted. Table 1 displays a two-by-two contingency table arraying actions (observed or not observed) and consequences (observed or not observed). The result is four archetypal information environments.

8Is this realistic for voters? As has been well understood by political scientists since the early 1960s, engaged interested voters usually have political preferences like this, but some unengaged relatively alienated voters do not ([Converse 1964], [Brookman 2016]). Such voters may not be very effective in selecting ideologically congruent politicians – though what congruence means when the voter-employer’s preferences are inchoate or incoherent is a tricky question.
Table 1: Archetypal Informational Environments in the Political Theory of Agency

<table>
<thead>
<tr>
<th>Actions</th>
<th>Outcome/Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observable</td>
<td>Observable</td>
</tr>
<tr>
<td></td>
<td>Full Transparency (FT)</td>
</tr>
<tr>
<td>Not Observable</td>
<td>Not Observable</td>
</tr>
<tr>
<td></td>
<td>Non-Transparent Consequences (NC)</td>
</tr>
<tr>
<td></td>
<td>Non-Transparent Actions (NA)</td>
</tr>
<tr>
<td></td>
<td>No Transparency (NT)</td>
</tr>
</tbody>
</table>

These are named following a taxonomy due to Fox and van Weelden 2012.

Under full transparency (FT), the voter can see both the incumbent’s actions and the resulting consequences. This situation would seem most favorable for control in a moral hazard setting and selection is an adverse selection one. However, when actions translate only probabilistically into outcomes, voters may still face a tricky inference problem (I see what he did, and I know it turned out fine, but was this because he was skillful? Or was he just lucky?)

At the opposite extreme, under no transparency (NT) the voter can see neither actions nor consequences. Clearly, this is the situation most ripe for moral hazard and adverse selection problems. As a practical matter, this environment is probably the most relevant baseline for most voters most of the time, and for some voters all of the time. Here, the voter will presumably have to rely on her presuppositions and presumptions about the incumbent and challenger.

An interesting environment is non-transparent action (NA) where the voter can observe outcomes but not actions. In the gardening analogy, imagine the employer leaves town for a while, so she can’t see what the gardener is doing. When she returns, suppose she finds a garden of thriving plants. If so, she is apt to infer that the gardener was skillful and hard-working in her absence. If she returns to a garden of blighted and dying vegetation, she is likely to infer that the gardener was unskillful, lazy, or both. The inference may be wrong in either case, for the garden may have flourished despite neglect or perished despite scrupulous attention. But if the garden is more likely to do well when the gardener is skilled and work hard, and more likely to do badly when he doesn’t, then the owner’s impression of the gardener – and her inclination to retain him – will be positively associated with the condition of the garden.

The typology is analytically equivalent to Wilson’s (1989) famous 2x2 typology of bureaucratic agencies.

More formally, a garden found in good condition will leave the owner with a more favorable impression of the gardener’s skill than one found in a worse condition if and only if the probability density \( f(\text{condition of garden}|\text{skill of gardener}) \) obeys the monotone likelihood ratio property (MLRP). See Milgrom 1981.
In a political setting, the voter can’t necessarily observe the actions of the president, but can see some evidence about the economy. Hence, the NA case is applicable to “classic” retrospective voting with respect to economic performance. Similarly, city voters can’t really see the president at work as manager but can see if it takes weeks to remove snow from the streets after a storm (as it famously did, with fatal electoral consequences, for Chicago Mayor Jane Byrne). In the case of the economy, many factors affect unemployment, inflation, growth, and trade, perhaps more strongly than most actions a president can take. But if the economy is somewhat more likely to do better when the president takes the right actions and somewhat more likely to do worse when he takes the wrong ones, rewarding or punishing the president based on the economy’s performance is sensible on the part of the voter, other things being equal.\footnote{In recent work, Chris Achen and Larry Bartels claim that in the NA information environment, voters sometimes punish presidents for events far outside their control – the so-call “shark attacks” critique of retrospective voting ([Achen and Bartels 2017]). Their empirical claims about the specific case of New Jersey shark attacks has collapsed under close scrutiny ([Fowler and Hall 2018]). However, recent work finds that U.S. governors have little or no impact on many conditions of concern to state voters, at least within electoral time horizons ([Dynes and Holbein 2020]). If so, tying voting to those conditions would make little sense for state voters. The scope and efficacy of retrospective voting in the NA environment remains open as an empirical question.}

Another well-studied environment is non-transparent consequences (NC). Here, the voter can observe the politician’s action but cannot observe the consequences of the action, at least before the election. So, in the gardening example, you can see the gardener watering or, seemingly judiciously, not watering. But it will take a long time for the plants to grow and you have to retain or fire the gardener today. Famously, this situation can lead to ”pandering” in which the politician takes a visible action the voter believes is correct – the popular action – even though the politician may know full well the ultimate result will be bad([Canes-Wrone et al 2001], [Maskin and Tirole 2004]). Conversely, the politician may refuse to take an unpopular action he knows is correct, because the good effects may come too late to save him at the election. For these reasons, political economist Andrea Prat calls accountability based purely on observed actions “the wrong kind of accountability” ([Prat 2005], see also [Fox and van Weelden 2012]).

### 2.2 Sentinels as Third-Party Information Providers

We now have enough pieces in hand to consider how a sentinel may affect the electoral accountability game. Following Arnold, a sentinel is a player who sees and reports 1) the politician’s action, 2) the
consequences of her action, or 3) the state of the world (when there are state-contingent utilities) at the time the incumbent took her action. A truthful report about these matters, if believed, has the effect of moving the voter from one information environment to another. So, using Table I, a believable report about the official’s action may shift the voter from NT to NC (e.g., “The President raised tariffs” or “The President abused his power.”) It may shift the voter from NT to NA (e.g., “The deficit is simply enormous.”) A believable report about the state of the world may shift the voter from NC to FT (e.g., “The President said he had to invade Iraq because of weapons of mass destruction, but it turns out there were no weapons of mass destruction.”) Or a credible report about both actions and consequences may move the voter from NT to FT (“The president raised tariffs and the resulting trade war devastated farmers.”) Depending on the ultimate information environment, the threat of the report may well induce an official to “do the right thing.” But as we note above, it may also induce her to pander or otherwise take actions a fully informed voter might prefer she not take.

The point of a TPA analysis such as the one we conduct below is to consider how the introduction of a sentinel might affect accountability under different circumstances, and then evaluate the consequences for voter welfare. But before doing that, we need to answer two preliminary questions: what incentives motivate different sentinels to make reports, and what factors make sentinel reports credible for the voter? To get traction here, we distinguish three kinds of sentinels: neutral conduits, interested parties, and challengers.

A neutral conduit simply reports the “truth, the whole truth, and nothing but the truth” as it knows it. A neutral conduit is forthcoming and candid – if it knows something, it will report it truthfully, and cover up nothing. And it is disinterested – it has no stake in the information itself or what the voter does with it. Neutral conduits approximate one ideal for the press, and for scientific experts. A strong intuition is, when neutral conduits exist, they will be very valuable to the voter.

In contrast, an interested party has a definite interest that it may hold irrespective of the actual state of the world, and potentially even irrespective of the identity of the incumbent officeholder. Suppose, for example, that the voter has state contingent preferences about infrastructure expenditures, of the following sort. The voter reasons, ”If our national infrastructure is run down (State 1), then I favor a big infrastructure plan even if taxes have to go up. But if our national infrastructure
is basically in fine shape (State 2) then I favor no plan and low taxes.” So the voter wants the infrastructure plan matched to the state of the world. A neutral conduit would relay any available information to the voter about the state of the world that prevailed in the official’s first term, and what the official did about it. But what will a concrete manufacturer or the association of civil engineers do? These actors are interested parties because they always favor a big infrastructure plan regardless of the actual state of our national infrastructure. So what can the voter infer from reports from civil engineers about the fitness of our national infrastructure? First, note that these parties will never pass on good news about infrastructure that would have motivated the incumbent to favor a small investment in infrastructure. The civil engineers association will always give our bridges, roads, and airports a failing grade ([Society for Civil Engineers 2017]). This means that silence from an interested party may actually connote good news. Now suppose the voters receive the anticipated message from the interested party: “It’s State 1 in America – our infrastructure’s a wreck!” What’s a voter to believe? If the information in the report comes verified by a trustworthy source or is nearly costlessly verifiable by the voter himself – that is, if it is hard information – then the report is informative. But if the information is unverified and unverifiable, the voter should be skeptical. A report that always comes out the same way regardless of the true circumstances doesn’t supply useful information about the state of the world; voters should disregard it. ([Milgrom 2008]).

Challengers are distinct from both neutral conduits and interested parties. As we discuss below, the key feature for the challenger is that he is in a zero-sum situation with the incumbent. Only one of the two contenders can win the election. Therefore, any information that hurts the incumbent is good for the challenger, and any that helps the challenger hurts the incumbent. Consider the infrastructure example again, and assume that reports from the challenger are nearly costlessly verifiable. Suppose the incumbent has correctly matched the state, supporting the infrastructure plan if it was State 1 but opposing it if it was State 2. The challenger will not want to point this out to voters, as doing so will help the incumbent – it makes the incumbent look skillful. The challenger will therefore remain silent. However, if the incumbent failed to correctly match the state, the challenger will delight in pointing out the incumbent’s blunder and apparent incompetence. The challenger’s motives are distinct from those of the interested party because challengers are happy to make any report about the state of world, so long as it hurts the incumbent; and stay silent,
regardless of the information available to them, when that information would help the incumbent.

3 Formalizing the Intuition

3.1 Preliminaries

To explore how the introduction of a sentinel can affect democratic performance – perhaps for the better, perhaps for the worse – we describe a highly stylized model. We begin with two players, an incumbent (she) and a voter (they), and then add a third: the sentinel (he). Drawing on the discussion in the previous section: the setting involves hidden information about the incumbent but in some scenarios also involves hidden actions as well; the model employs state contingent preferences; and the key hidden information about the incumbent is her skill in discerning the state of the world.

There is a state of the world, $\omega$, which can take on values of 0 or 1. Both states are equally likely, which is common knowledge. An incumbent can be described by her “type,” $t$, which can be either low ($L$) or high ($H$). The incumbent initially knows her own type, but the voter has some uncertainty about the incumbent. Specifically, from the voter’s perspective at the beginning of the game, the prior probability the incumbent is type $H$ is given by $\alpha_i$, and the probability that the incumbent is type $L$ is $1 - \alpha$. The voter has a corresponding belief about the challenger’s type, described by the parameter $\alpha_c$. Both $\alpha_i$ and $\alpha_c$ lie between zero and one.

The incumbent receives a signal $\theta$ about the state of the world, which can take on values of either 0 or 1. The signal $\theta$ is correlated with the state of the world $\omega$, but the quality of the signal depends on the incumbent’s type. Specifically, high-type incumbents receive perfect signals about the state of the world ($\Pr(\theta = \omega | t_i = H) = 1$). In other words, if the state of the world is 1, a high quality incumbent will know it’s 1 for sure; if 0, she will know it’s 0. Low-type incumbents, by contrast, receive imperfect signals. Specifically, their signals are correct with probability $q$ (i.e., $\Pr(\theta = 1 | \omega = 1, t_i = L) = \Pr(\theta = 0 | \omega = 0, t_i = L) = q$), where $q$ lies between $\frac{1}{2}$ and 1. We will often refer to $q$ as the quality of the low-quality incumbent’s signal.

The incumbent must take a policy action $a \in \{0, 1\}$. In other words, she must choose between two policy actions, though one can interpret $a = 0$ as ”do nothing” or ”retain the status quo.” The incumbent’s action translates into consequences for voters in a simple way: if the incumbent state-matches then the result is good for the voters. But if the incumbent fails to state-match, the
result is bad for the voters. So, one might think of the voter as having a utility function of the form:

\[ u_v(a, \omega) = \begin{cases} 
1 & \text{if } a = \omega \\
0 & \text{otherwise}
\end{cases} \]

In the interest of simplicity, we employ the following device. Let \( \tilde{\alpha}_i(I) \) denote the voter’s posterior probability that the incumbent is high quality given information \( I \). We assume the probability the voter retains the incumbent is equal to \( F(\tilde{\alpha}_i(I) - \alpha_c) \), where \( F(\cdot) \) is a smooth, strictly increasing function bounded between zero and one. This formulation is intended to capture the fact that while the incumbent benefits, ceteris paribus, from voters’ positive impressions of her quality, other (stochastic) features of the political environment, as well as her impression of the challenger, may also affect the voter’s ultimate selection of candidates. The key element in the formulation is the voter’s posterior beliefs about the incumbent, which must be Bayesian rational wherever possible.

We assume the incumbent wishes to maximize the probability she retains office. So, the incumbent (or her ally) will seek to maximize \( \tilde{\alpha}_i(I) \), whereas an opponent (possibly a challenger, but also potentially a non-aligned media outlet) will seek to minimize it. The zero-sum nature of the competition between the incumbent and the challenger is a key point.

Setting up the model in this fashion “rolls up” play in the second period of an archetypal TPA game into the first period payoffs (a move noted in Section 2). In the one period game, in the absence of a sentinel, the incumbent faces an essentially decision-theoretic problem, though one strongly shaped by the voter’s rational updating of beliefs. The addition of a sentinel then creates a game between the incumbent and the sentinel. Their strategic interactions shape voter posterior beliefs and hence incumbent and challenger payoffs through the re-election function. This way of setting up the TPA game greatly simplifies the analysis and allows us to focus on the essential elements of fire-alarm accountability.

As discussed in Section 2, both incentive/sanctioning effects and politician selection effects are critical components of a theory of democratic accountability. Our model allows us to study both. A critical question we ask is the following: can “virtuous behavior” by the incumbent benefiting the voter be sustained in equilibrium?\(^{12}\)

\(^{12}\)This does not necessarily imply or require that virtuous behavior be the unique equilibrium.
virtuously is simply one who follows her signal. To see why, note that clearly, when the high-quality incumbent follows her signal, she assuredly state-matches since her high quality signal is correct with 100% probability. Hence, following the signal definitely leads to good consequences for the voter. But since we assumed both states of the world are equally likely to begin with, and, further, that the low-quality incumbent’s signal is right more often than not (even if not especially often), when the low-quality incumbent follows her signal she maximizes the expected benefit to the voter as well.

The second critical component in the theory of democratic accountability concerns how much voters learn about the incumbent’s quality, thus permitting them to make more informed choices at election time. So the issue is, in equilibrium do voters learn much about the incumbent’s type? While our simple model has only one period, the probabilistic vote function effectively captures a more complicated, multi-period model in which voters enjoy downstream benefits from having a high-type incumbent in office. These benefits are more likely to accrue to voters, *ceteris paribus*, when they have better information about the incumbent’s type, as it will permit them to make fewer errors in determining whether the incumbent or challenger is the better choice.

We now turn to the four informational environments in which voters might find themselves in the absence of a sentinel. The four information environments were indicated in Table 1.

**Voter observes neither policy action nor outcome.** This is the NT (No Transparency) scenario in Table 1. This information environment is the simplest case. If the voter can observe neither the incumbent’s action \( a \) or whether it was correct (i.e., whether \( a = \omega \)), there is nothing that the incumbent can do to alter the voter’s beliefs. The voter will then reelect with probability \( F(\alpha_i - \alpha_c) \). Note that because the incumbent’s electoral fortunes are unaffected by her policy choice, any policy choice in any incumbent information set is an equilibrium. Of course, this includes “doing the right thing” by following the signal. We can easily break this indifference by giving the incumbent some infinitesimal benefit from pursuing the voter’s interests. In that case, behaving virtuously is the unique equilibrium. But voters still won’t learn about the incumbent’s type from that behavior.

**Voter observes policy action, but not outcome.** This is the NC (No Consequences) scenario, for example, in one of our baseline examples, any behavior by the incumbent, including virtuous behavior, can be sustained as an equilibrium.
scenario in Table 1. Such an information environment is almost as simple as the NT baseline. To see why, suppose that in equilibrium, both types of incumbent behave virtuously. Because we have deliberately set up the model so that both states of the world are equally likely, and so that the accuracy of the incumbent’s information is independent of the state, the voter will be just as likely to observe $a = 0$ as $a = 1$. Hence, the policy will provide no new information about the incumbent’s type. And given this, the incumbent has no incentive to deviate from virtuous behavior – just like in the case in which the voter observes neither the policy action nor the outcome.

Note that this virtuous behavior is sustainable because of a deliberate modeling choice on our part – making both states of the world equally likely. This eliminates any incentive of the incumbent to “pander” to voters by choosing the policy more consistent with their prior (cf., [Canes-Wrone et al 2001]).

Voter observes the outcome, but not the policy action. This is the NA (No Actions) scenario in Table 1. In this informational environment, the voter learns whether the incumbent was right or wrong in their policy choice, even though she cannot observe incumbent actions. Now suppose both types of incumbent behave virtuously. High-quality incumbents will always choose the correct policy, and low-quality incumbents will choose the correct policy more often than not. Given that the incumbent behaves virtuously, the voter, upon learning that the wrong policy was chosen, will know with certainty that the incumbent is of low quality ($\tilde{\alpha}_i(a \neq \omega) = 0$), thus decreasing the probability the incumbent is retained. If, by contrast, the voter learns that the correct policy was chosen, then the voter will know that it was chosen either by the high quality incumbent or a low quality incumbent who received a signal that turned out to be correct. The voter’s posterior belief on the incumbent will then (by Bayes’ Rule) be equal to, $\tilde{\alpha}_i(a = \omega) = \frac{\alpha_i}{\alpha_i + (1 - \alpha_i)q}$, which is strictly greater than $\alpha_i$, thus increasing the probability the incumbent is retained. Given the foregoing, the low-quality incumbent has every incentive to get the policy right, which she maximizes the probability of accomplishing by behaving virtuously.

Voter observes both the outcome and the policy action. This is the FT (Full Transparency) scenario in Table 1. Given the setup of the model, the logic is identical to the previous case.

We summarize the no-sentinel baselines in Table 2, which mirrors Table 1. It is worth noting, in terms of motivating virtuous actions, our simplifying assumptions bias the model in favor of the
incumbent doing right by the voter. This allows a clear baseline from which to consider whether the strategic game between the incumbent and the sentinel actually improves voter welfare.

### 3.2 Introducing the Sentinel

Our sentinel-free analysis, summarized in Table 2, demonstrates how access to verifiable information about the outcome can enhance accountability, both by strengthening the incentives for the incumbent to behave virtuously, and by enhancing the ability of voters to select good types. We also see the limitations of informational environments in which the voter can observe the incumbent’s action but not the consequences of the choice.

With these considerations in mind, consider the following permutation of the model. Suppose we are in a world in which absent any fire alarm, the voter observes neither the policy nor the outcome (the NT information environment). A third party, whom we call the sentinel, receives some information that he may pass on to the voter. Specifically, if the incumbent chooses policy action $a = 1$, the sentinel receives evidence that the policy was either right (if $\omega = 1$) or wrong (if $\omega = 0$) with probability $\pi_1$; with probability $1 - \pi_1$, the sentinel receives no such information. Likewise, if the incumbent chooses policy action $a = 0$, the sentinel receives evidence that the policy was either right or wrong with probability $\pi_0$, and with probability $1 - \pi_0$ receives no such information. Note that $\pi_1$ need not equal $\pi_0$.\(^{13}\)

13This setup resembles the “asymmetric resolution” extension in [Canes-Wrone et al 2001].

There are now several things to consider. First, consider the preferences of the sentinel, as discussed in Section 2:

- The sentinel may be a neutral conduit, sharing the voter’s preference for a high quality incumbent;
- The sentinel may be a challenger, who benefits when the incumbent’s reputation suffers; or

<table>
<thead>
<tr>
<th>Actions</th>
<th>Observable</th>
<th>Not Observable</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT: virtuous equilibrium</td>
<td>unique, some updating</td>
<td>NC: virtuous equilibrium possible, no updating</td>
</tr>
<tr>
<td>NA: virtuous equilibrium</td>
<td>unique, some updating</td>
<td>NT: virtuous equilibrium possible, no updating</td>
</tr>
</tbody>
</table>

---

\(^{13}\)This setup resembles the “asymmetric resolution” extension in [Canes-Wrone et al 2001].
- The sentinel may be an *interested party* that is biased in favor of one policy (say, $a = 0$), irrespective of the actual state of the world.

Obviously, in a world of self-interested political actors, we would have strong reason to believe that the sentinel is unlikely to be a neutral conduit. Nonetheless, a neutral and honest sentinel is a useful benchmark against which to compare the challenger and activist sentinels.

Next, consider the nature of the evidence the sentinel may pass on to the voter. Two extreme cases are the following:

- The information may be *cheap talk*, in the sense that it may or may not be taken as true; or
  - the information may be *hard*, in the sense that it will be taken as fact by a voter presented with it.

Of course, may other kinds of evidence are possible. Information may be verifiable at cost, or verifiable at cost with some probability, for example. To convey the intuition as expeditiously as possible, we will abstract away from these cases and consider only the more stark examples above.

### 3.3 The Cheap Talk Fire Alarm

Our first step in demonstrating the contingency of the purported democracy-enhancing properties of fire alarms is to consider them when evidence is cheap talk. Under such conditions, for informative communication between the sentinel and the voter to be possible, the latter must take the former’s word as given.

**Neutral Sentinels with Cheap Talk.** First, consider the neutral sentinel, who is known to be neutral by the voter. By construction, this actor shares the preferences of the voter, and is thus, in a sense the perfect agent. Because of the common interest in high-quality incumbents, an equilibrium exists in which the sentinel neutrally and honestly conveys information to the voter (when he has it) and the voter believes the sentinel and acts accordingly.

Moreover, consider the incentive effects of the neutral sentinel on the reelection-minded incumbent. Recall that high-quality incumbents can always choose the policy that is correct from the voter’s perspective. Suppose she does so. Then the sentinel will either have good news to pass on to the voter (the policy was correct), or no news. By contrast, the low quality incumbent may
err, generating bad news that the sentinel will dutifully, and credibly, pass on. Bad news perfectly reveals that the incumbent is a low-type. Clearly, then, low-quality incumbents have an incentive to minimize the probability of bad news. They maximize the probability of doing this by behaving virtuously.

If we understand “fire alarms” to be signals purporting malfeasance (bad news), then in the presence of a neutral sentinel, there is an equilibrium in which fire alarms (1) are taken seriously by voters; (2) strengthen the incentives of the incumbent to behave virtuously; and (3) help voters make more informed choices come election time. Note that this intuition would be preserved even if the only evidence available to the neutral sentinel were evidence of bad news – i.e., “true” fire alarms. Incumbents would still be motivated to minimize the probability of bad news, which they accomplish by behaving virtuously. This situation seems to confirm Arnold’s optimism about sentinels in an electoral context.

**Challenger Sentinels with Cheap Talk.** Of course, in politics, neutral conduits may be hard to come by. As noted in the Arnold quotation in the Introduction, the sentinel is much more likely to be a self-interested actor like a challenger or activist, whose interests may not perfectly align with those of the voter. It turns out that if talk is cheap, there exists no informative equilibrium in which the voter takes statements by the sentinel as credible.

It is easy to see why. If the sentinel is a challenger, his objective is to damage the incumbent’s reputation. Now suppose, again, that high-quality incumbents always choose the policy that matches the state of the world. Irrespective of the information actually received by the sentinel, he will always have an incentive to try to convince the voter that he received information about incumbent performance and that it was bad – i.e., that the policy chosen did not match the state. But because this motivation persists regardless of the truth, the voter won’t believe him, and will learn nothing from his utterances. Not all is lost, however: the incumbent will still weakly prefer to pursue the policy expected to most benefit the voter, just as she did in the case with no sentinel.

**Interested-Party Sentinels with Cheap Talk.** As the passage from Arnold notes, sometimes the sentinel is an activist or advocacy group. Are things better when the sentinel is an interested party? Not really. Unlike the voter (and the challenger), the activist does not care about the occupant of the office. He is happy as long as his favored policy is pursued. Consequently, any pronouncements he makes will be independent of the actual state of the world. For example,
he may say “My favored policy was chosen, and moreover it was the correct choice!” Or, “The incumbent chose the policy I hate, and boy, did it turn out badly.” In any case, the voter will not find these statements credible and again, will learn nothing. But also again, the incumbent will continue to at least weakly prefer to do right by the voter.

The upshot of the foregoing is that if information is cheap talk, the possibility that fire alarms will improve the voter’s lot compared to a counterfactual world with no fire alarms is contingent on very specific assumptions about the preferences of the sentinel pulling the alarm.

3.4 Fire Alarms with Hard Evidence

A reader might, given the foregoing, be tempted to jump to one of two conclusions: either fire alarms only benefit the voter under rare circumstances (the truly neutral and honest sentinel), or the failure of fire-alarms to help voters is a mere artifact of the artificial cheap talk environment in which we have considered them. This conclusion would be premature at this point, however. We first need to consider how things work in a different informational environment. Here we do so by supposing that information is hard rather than cheap talk. In other words, we consider a setting in which, if the sentinel receives “good” news that the policy correctly matched the state, that evidence is both accurate and dispositive: thus, if the sentinel provides the news to the voter, she will accept it at face value. Likewise, if the sentinel has “bad” news and transmits it to the voter, the voter will accept it as true. In such a situation, we will also need to consider what voters will believe if they receive no news. No news admits two possibilities: either the sentinel had no information to transmit; or, the sentinel had information but suppressed it – a cover-up.

Neutral Sentinels with Hard Information. In this stark informational environment, consider first the neutral-conduit sentinel. In the cheap talk setting, this actor conveys his possession of good news, bad news, or no news, and the credibility of these signals is established endogenously in equilibrium given the likemindedness of the voter and sentinel. In other words, “the truth will out,” at least to the extent that there is truth available to the sentinel to reveal.

Moving away from cheap talk to a hard information environment changes nothing: the sentinel will still report information available to him, and now, the evidence will be even firmer than it was in a situation where it was already accepted at face value. Consequently, the truth will continue to out. Voters will have access to all available information, facilitating their goal of selecting the
best available officeholder. And incumbents, for their part, will have strong incentives to behave virtuously, as doing so maximizes their probability of reelection.

**Challenger Sentinels with Hard Information.** In a world where high quality incumbents always choose the correct policy, the challenger has a clear incentive to report all bad news regarding the incumbent’s choice to the voter, because the voter will infer from bad news that the incumbent was a low type, thus enhancing the challenger’s electoral prospects. Just as bad news undermines the incumbent’s reputation in the eyes of the voter, so too does good news enhance it; hence, the challenger will suppress any good news it has at its disposal. If a challenger sentinel reports all bad news and suppresses all good news, and voters interpret bad news as definitively establishing the incumbent is a low-type, then a high-type incumbent will clearly have an interest in selecting the correct policy: doing so guarantees no news, whereas deviating and choosing the wrong policy will result in a lottery between the electoral consequences of no news and bad news. No news is better.

But now consider the problem from the perspective of the low quality incumbent: she has an incentive to minimize the likelihood of bad news. This was also the case when the low quality incumbent faced a neutral sentinel, of course. But given the special incentives of the challenger, does this situation now mean a departure from virtuous behavior by the incumbent? It turns out that the answer is yes, at least under some conditions.

Specifically, suppose $\pi_1 > \pi_0$, so that the challenger is more likely to receive hard information to pass on to voters when the incumbent has chosen $a = 1$ (e.g., depart from the status quo) than when she has chosen $a = 0$ (e.g., maintain the status quo). If the incumbent is a low type, and receives a signal of $\theta = 0$, she has two reasons to follow her signal and choose the corresponding action of $a = 0$, to the benefit of the voter: the choice is more likely to be correct, and if it is incorrect, it is less likely to yield bad evidence for the challenger to transmit to voters.

If the incumbent is low quality and receives a signal of $\theta = 1$, however, she faces a tradeoff given the threat of a fire alarm: if she chooses $a = 1$, she is more likely to be correct, but if she is incorrect, the challenger will “have the goods” on her with relatively high probability. If, on the other hand, she chooses $a = 0$, she is less likely to be correct, but if she chose incorrectly, it is less likely that the challenger will actually receive the bad news to pass on to the voters. It turns out that given a signal $\theta = 1$, the low quality incumbent will take the non-virtuous action that *hurts*
the voter \((a = 0)\) in expectation if and only if

\[
\frac{\pi_1}{\pi_0} > \frac{q}{1 - q}.
\]

In other words, if the risk of bad news associated with choosing the policy that is correct in expectation is sufficiently great relative to the low quality incumbent’s expertise, the low-quality incumbent will choose the policy more likely than not to be wrong.

The upshot of the foregoing analysis is that the presence of the sentinel threatens to create a distortion in the incentives of (low-type) incumbents relative to a world with no sentinel. Note, however, that the overall effect of the sentinel’s presence on the voter’s well-being is ambiguous. This is because relative to the baseline with no sentinel, voters can update their beliefs about the incumbent’s type: those beliefs will be revised downward (to zero, in fact) given the provision of bad news; and they will be revised upward given the provision of no news. Hence, our partially-informed voter is more likely to get a high-type incumbent in office after the election. Thus, the incentive and selection effects from a challenger sentinel are in tension, making overall conclusions regarding voter welfare ambiguous in the absence of stronger assumptions.

**Interested-Party Sentinels with Hard Information.** Next, suppose the sentinel is an interested party – say an activist or advocacy group – and in particular, one who cares only that the incumbent selects the policy action \(a = 0\), irrespective of the state of the world. With these preferences, the group is indifferent with respect to the occupant of the office, and can craft a fire alarm strategy that maximizes the incumbent’s incentive to choose \(a = 0\), even when she receives the signal suggesting she ought take the opposite course \((\theta = 1)\). Consider the following strategy for the sentinel: report available good news if and only if the incumbent turns out to have correctly chosen the action \(a = 0\); and report available bad news if and only if the incumbent turns out to have incorrectly chosen \(a = 1\). Given this strategy from the sentinel, a high quality incumbent who always follows her signal need never fear bad news – from her perspective, no news is the worst possible outcome. And so, the voter could again infer from bad news that the incumbent was low quality with certainty.

Now consider the game from the perspective of the low quality incumbent. She will, of course, prefer to follow her signal given \(\theta = 0\) – doing so maximizes the odds that the voter will receive
good news about her performance. The interesting question concerns what she will do when \( \theta = 1 \), which could place her at odds with the interested-party sentinel. Suppose she behaved virtuously in that circumstance and chose \( a = 1 \). Then with probability \((1 - q)\pi_1\), she would be wrong, and the advocacy group would learn about it and publicize the bad news to the voter, driving the incumbent’s reputation \( \tilde{\alpha}_i \) down to zero. With complementary probability \((1 - (1 - q)\pi_1)\), the voter would receive no news. But recall that from the voter’s perspective, receiving no news is also consistent with the incumbent being high quality. Clearly, the low quality incumbent would prefer no news to bad news.

Given this preference, the low quality incumbent can improve her lot by deviating from virtuous behavior by choosing policy \( \alpha = 0 \) when her signal \( \theta = 1 \). With probability \((1 - q)\pi_0\), this was actually the right move and the advocacy group will have good news to convey to the voter about the incumbent’s correct choice. With complementary probability, there will either be no news to report, or bad news that the interested-party sentinel will suppress – hence, no news from the perspective of the voter. But a lottery between good news and no news is clearly better from the low-quality incumbent’s perspective than a lottery between bad news and no news. Hence, virtuous behavior by the low quality incumbent given \( \theta = 1 \) cannot be an equilibrium. Given reasonable restrictions on beliefs off the path of play (namely, if we assume the voter will infer the incumbent is low quality given bad news), we can establish the converse: the low quality incumbent’s disregarding the signal \( \theta = 1 \) is consistent with equilibrium play. As in the case of the challenger sentinel, the activist sentinel induces incumbents sometimes to do the wrong thing from the voter’s perspective. Note that unlike in the case of the challenger sentinel, this will be the case irrespective of the underlying parameters \((q, \pi_0, \text{and } \pi_1)\).

Despite the bias in reporting from an interested-party sentinel, voters will, on occasion, have access to hard information about the incumbent’s performance (although, in equilibrium, the voter will never receive any bad news due to the distortion in the low quality incumbent’s behavior). Accordingly, the voter’s selection problem will be mitigated relative to a baseline in which the voter has no access to information about policy or performance. As in the case with the challenger sentinel, however, there is again a trade-off from the voter’s perspective between the selection benefits and the incentives for some incumbents to do the wrong thing, some of the time.
3.5 Can Incumbent Credit-Claiming Mitigate Biased Fire Alarms?

Up to this point, we have considered cases in which the sentinel is the only source of information available to the voter. Of course, we know that incumbents are generally eager to claim credit for any good news, whether or not it is actually associated with their actions in office (Mayhew 1974). A natural question to ask, then, is whether extending the model to permit the incumbent to transmit good news to the voter about their performance might mitigate some of the distortions associated with biased sentinels that we described in the previous section. Needless to say, if communication is pure cheap talk, there can be no credible communication between the re-election-minded incumbent and the voter: the former will always want to convince the latter she is a high-type, irrespective of her actual type.

Accordingly, we restrict our attention in this section to the hard information environment. To the extent that “the truth will out” given a neutral sentinel, the information conveyed to the voter via credit-claiming will be redundant. Accordingly we would anticipate no changes to the welfare of the voter – either through the selection or incentive channels – given the presence of a credit-claiming incumbent and neutral sentinel.

Consider, next, the case of the challenger sentinel. Recall that the challenger transmits information to the voter if and only if it is bad news for the incumbent. By the same logic, the incumbent will only transmit information to the voter if it is good news (remember – if high quality incumbents have the ability to select the correct policy with no error, then bad news will fully reveal that the incumbent is low quality). So with the addition of a credit-claiming incumbent and a detractor challenger, all available evidence will ultimately be passed on to the voter: good news by the incumbent, and bad news by the challenger. But then, from the perspective of the voter, we are in a world that is equivalent to one with the neutral sentinel, which we have already established is excellent from the voter’s perspective: the selection problem is mitigated because of the rich information available to the voter; and there are no distortions in the low quality incumbent’s incentives. Interestingly, this situation closely resembles an idealized view of an adversarial legal system, in which prosecutors have incentives to present all available incriminating evidence to a jury, while defendants have incentives to present all available exculpatory evidence (Dewatripont and Tirole 1999).
The question of whether credit-claiming by the incumbent can mitigate the biased information coming from an interested-party sentinel is more subtle. To be sure, when we allow for credit-claiming with hard information, more information can potentially reach the voter. Specifically, suppose a low-type incumbent were to choose the policy disfavored by the activist. In the absence of credit-claiming, the voter would only observe no news or bad news. Add a credit-claiming incumbent and now good news is also a possibility. More information will be revealed than before, and the downside electoral risk of crossing the interested-party sentinel will be muted.

Suppose, by contrast, the incumbent chooses the policy the activist wants despite receiving the signal suggesting the contrary action. Now, both the incumbent and the activist have a mutual interest in suppressing any bad news about that policy. By catering to the activist’s preferences to the detriment of the voter, the worst outcome for the incumbent – bad news – can be avoided. Implicitly, the incumbent colludes with the interested party, against the voters.

The downside risk of bad news to the low quality incumbent may thus continue to motivate her to insure against its adverse consequences by selecting the policy favored by the interested-party sentinel, even if she believes it is the wrong policy. However, this downside risk is offset to some extent by the potential upside: credit-claiming if she chooses the policy the advocacy group dislikes and it turns out to have been the right choice (“standing up to the special interests”).

Given the foregoing analysis, introducing hard information credit-claiming doesn’t bring us quite back to a situation equivalent to that of the neutral sentinel, as it did in the case of the challenger sentinel. In supplementary analysis, we demonstrate the existence of conditions under which virtuous behavior cannot be sustained in equilibrium, even given the otherwise ameliorating effects of credit-claiming. In other words, the distorting effects of a biased sentinel cannot be fully eliminated when the sentinel is an interested party. Here are some observations concerning the conditions under which the incumbent will succumb to the activist’s demands (choosing $a = 0$ even given a signal of $\theta = 1$) to the detriment of the voter:

First, unsurprisingly, the low quality incumbent will be more inclined to implicitly collude when $q$, the accuracy of her own information, is relatively low, because the risk that the interested-party sentinel will publicize bad news if she sticks with her signal and chooses $a = 1$ will be correspondingly high.

Second, consider that in previous cases, the incumbent’s choice was relatively simple: she
was, for example, choosing between lotteries of no news and bad news, or comparing a lottery between no news and good news with a lottery between no news and bad news. By contrast, with an interested-party sentinel and hard information credit-claiming, she must compare a lottery in which no news, good news, and bad news are all possible (occurring when she deviates from the sentinel’s preferred policy) with one in which only no news and good news are possible (when she caters to the sentinel). Which option dominates will depend on the relative value of good news and no news, which is encapsulated in the ratio of the voter’s posterior beliefs under those two circumstances: $\tilde{\alpha}_i(\text{good news})/\tilde{\alpha}_i(\text{no news})$. This ratio is decreasing in the voter’s prior belief about the incumbent, $\alpha_i$, because the upside potential of good news is lower when the voter already believes that the incumbent is high quality. Hence, an incumbent who is actually low quality when the voters are inclined to believe she is high quality will be most predisposed to “play it safe” by implicitly colluding with the activist group.

### 3.6 Summary

The implications derived from our simple model concerning selection and incentive effects are summarized in Table 3. Regarding the selection component of democratic accountability, introducing a sentinel cannot hurt (no revelation), and may possibly help (partial or full revelation), a voter distinguish high- and low-quality candidates for office. Some sentinels will strategically withhold information in pursuit of their own interests, but voters can adjust for this if they understand the sentinel’s motives. This serves as a partial confirmation of the conventional understanding of fire alarms as beneficial to a principal in a principal-agent relationship.

Regarding the incentive component of democratic accountability, however, our conclusions are less sanguine. Whereas the presence of sentinels won’t distort the incentives of the incumbent if sentinel communications are cheap talk, the presence of a sentinel may indeed distort incentives when sentinels can present hard information to voters. Distortions take the form of selecting a policy not because it is best ex ante from the voter’s perspective, but because it minimizes the probability of bad news. In essence, the incumbent may implicitly collude with the sentinel. Allowing the incumbent a channel to publicize her positive accomplishments (using hard information) can eliminate the distortion when the sentinel is a challenger, but not necessarily if he is an interested

---

14This assumes that the voter processes information rationally. If we relax this assumption – for example by introducing confirmation bias or violations of negative introspection, all bets are off.
Table 3: The Conditional Effects of Third-Party Sentinels on Democratic Accountability

<table>
<thead>
<tr>
<th>Sentinel Type</th>
<th>Cheap Talk Revelation</th>
<th>Distortion</th>
<th>Hard Information Revelation</th>
<th>Distortion</th>
<th>Hard Info + CC Revelation</th>
<th>Distortion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>Full</td>
<td>None</td>
<td>Full</td>
<td>None</td>
<td>Full</td>
<td>None</td>
</tr>
<tr>
<td>Challenger</td>
<td>None</td>
<td>None</td>
<td>Partial</td>
<td>Sometimes</td>
<td>Full</td>
<td>None</td>
</tr>
<tr>
<td>Activist</td>
<td>None</td>
<td>None</td>
<td>Partial</td>
<td>Always</td>
<td>Partial</td>
<td>Sometimes</td>
</tr>
</tbody>
</table>

Notes: (1) Credit-claiming abbreviated as CC; (2) Full revelation means full revelation of available evidence.

4 Example: The Informational Environment of Criminal Justice Policy

By the mid-1990s and early 2000s, a persistent feature of American public life was the odd juxtaposition of a marked decline (from a peak in around 1991) in crime rates with the enactment, at the local, state, and national levels, of ever-more punitive criminal justice policies. At the state level, provisions such as truth-in-sentencing laws (requiring convicted felons to serve the full term, or some minimal proportion, of the sentence received at trial), mandatory minimum sentences, the abolition or severe curtailment of parole, and sentencing enhancements for repeat offenders, all contributed to a sharp increase in the proportion of American citizens under some form of correctional control. This proportion, around 2.2 individuals per 1,000 in 1980, peaked at 7.6 per thousand from 2006 to 2008; it was 6.7 per thousand in 2016. Figure 1 captures this discrepancy, displaying crime and imprisonment rates over a nearly sixty-year span of U.S. history.

Although the rise in incarceration rates preceded it, several provisions of the 1994 Violent Crime Control and Law Enforcement Act exacerbated the increase. Among other things, the law dramatically increased penalties for federal crimes (including a “three strikes” provision for repeat offenders), and provided billions in grants to states to finance new prison construction conditional on their adoption of truth-in-sentencing laws.

Coincident with the rise in incarceration was an increased willingness at the local level to employ aggressive policing practices, such as stop, question, and frisk (SQF) and militarized police tactics such as SWAT deployment ([Mummolo 2018](https://www.bjs.gov/index.cfm?ty=kfdetail&iid=493)). The antecedents of these developments are

---


\[16\] According to the Bureau of Justice Statistics, a felon is defined as imprisoned if his or her sentence exceeds one year of confinement; we use imprisonment rates rather than incarceration rates owing to the availability of the longer time series.
numerous, but observers point to a number of policies that may contribute to them. At the state and local level, asset forfeiture laws and the reliance of localities on fines and fees as sources of revenue (Hye Young) create high-powered incentives to police aggressively, particularly in disadvantaged communities of color (DOJ Report). Federal law and policies adopted by federal agencies have also contributed in this regard: the 1994 Crime bill provided funds for hiring 100,000 new police officers. The drug war, combined with the cheap availability of surplus military hardware, have contributed to militarization ([Balko 2013]). And the Justice Department’s Equitable Sharing program (suspended under the Obama administration but reinstated under Trump) shares a portion of assets seized by federal law enforcement with state and local agencies, even in states that do not have asset forfeiture laws of their own.

Until recently, the conventional wisdom among observers of criminal justice politics and policy is the existence of a ratchet effect, wherein elected officials would pursue ever more punitive policies to reign in crime, irrespective of their social or financial costs. The logic underlying this argument is that whereas the cost of crime is easy to see, the cost of crime control is typically obscured from the public eye. In the language of our model, hard evidence that a criminal justice policy is too lenient is more likely to materialize than hard evidence that a policy is too stringent. And so, elected officials concerned with their own political survival might err on the side of being too punitive, to minimize the risk of “bad news” that challengers and advocacy groups (in this policy space, police unions and victims’ rights organizations) would have every incentive to publicize.

Critically, the political incentives to support ever more punitive policies, according to this account, do not depend on the ideology of the officials. And so, we saw large majorities of Democrats – and even large majorities of the Congressional Black Caucus (though see [Hinton et. al. 2016], who note serious reservations) – supporting the 1994 Crime Bill.

The experience of the last decade, however, has prompted a significant push toward decriminalization, decarceration, and less aggressive policing. In the 2016 Democratic primary, support for the 1994 crime bill proved a liability for then-candidate Hillary Clinton. Democratic politicians such as New York mayor Bill De Blasio have successfully run on pledges to significant reign in SQF. California has adopted sentencing reforms to to curb prison overcrowding. And Philadelphia, St. Louis, and a number of other jurisdictions have elected district attorneys committed to criminal justice reform.
What changed between the 1990s and the 2010s that loosened the incentives of Democratic politicians to support at least tentative reforms of the criminal justice system? A number of factors are clearly at play, but one is a change in the availability of hard information about the downsides of punitive criminal justice policy. Clearly, it has been increasingly difficult to ignore the effects of mass incarceration on communities of color. But technology has also played a significant role of bringing sustained attention to the adverse consequences of incarceration and aggressive policing, beyond those communities. A critical development is the ubiquity of cell phone cameras, which has brought instances of police-initiated violence into the public eye at a dramatically increased rate. In the language of the model, the probability that hard evidence of overpunishment is revealed has increased relative to the probability that underpunishment is revealed (which has, by contrast, remained relatively fixed). Accordingly, officials have reason to fear the electoral consequences of both kind of “bad news.” Indeed the political potency of groups such as Black Lives Matter rests in large part on the change in the informational environment.

5 Conclusion

As a discipline, political science has long been preoccupied with the implications of an electorate composed in large part of uninformed citizens. In coming to terms with this hard fact of political life, many have pointed to elites as an imperfect solution: by providing cues, elites may help “rationally ignorant” voters vote as if they were truly informed. In this paper, we have examined a subtly different role for elites in the political information business. Specifically, we have sought to clarify how third party “sentinels” with the ability to convey information about incumbent performance to voters in the form of “fire alarms” can serve to enhance or undermine democratic accountability. Our analysis reveals three critical lessons.

First, if the information conveyed by a sentinel is unverifiable cheap talk, then only sentinels understood to be neutral, unbiased conduits of information to voters can credibly communicate information about incumbent performance to voters. By contrast, fire alarms sounded by challengers who hope to tarnish the reputation of the incumbent, or interested parties who hope to burnish the reputation of their favored policy irrespective of its utility, will tend to be ignored by voters when those messages are not accompanied by hard evidence.

Second, if a sentinel’s message contains hard information, then the presence of a challenger
or interested party sentinel may serve to undermine democratic accountability, by convincing low quality incumbents to minimize bad news, even if doing so means undermining voter welfare. In the presence of these biased sentinels, there is an inherent tradeoff between these distortionary incentive effects, which hurt voters, and a richer informational environment, which helps them select good incumbents.

Third, the possibility of incumbent credit claiming can mitigate the distortions induced by biased sentinels, but only under some circumstances. Specifically, distortions will dissipate in the presence of a challenger who only reports bad news about incumbent performance to voters, as they can be countered by the strong incentive of the incumbent to report good news. But distortions may persist when the sentinel is an interested party such as an activist or advocacy group. This is because occasions will emerge in which a low quality incumbent can collude with the sentinel to suppress bad news about her performance. The price is the willingness, on occasion, of low quality incumbents to cater to the sentinel’s wishes even when it is not in the voter’s interest.

A natural question to ask concerns the possibility of fire alarm accountability in a polarized age. Are voters looking for competent representatives, or simply ideologically well-aligned ones? If the latter, then the limitations of the model we have described above are obvious. And yet, we have no reason to believe that in a different model, in which the uncertainty of the voter concerned whether the incumbent was a moderate or a true believer, all distortions would disappear. As noted above, developing such a model is a task we leave to future research. And yet, at the very least, we can think of our analysis as a cautionary note concerning what kinds of information third parties may bring to bear to help inform voters and create the right incentives for incumbents.

Finally, our model suggests several empirical applications concerning the potential distortions created by biased sentinels. Two that are particularly important concern the availability of challengers and interested parties to serve as sentinels, and the informational environment that governs the production of tangible evidence about policy performance. With respect to the former, we would anticipate fire alarm oversight provided by challenger sentinels to be reduced in safe districts. On the one hand, the standard intuition is that the absence of viable challengers would undermine the accountability of the incumbent. However, this deleterious effect may be mitigated by the removal of the distortionary fear of challengers as self-interested bearers of bad news. Likewise, when thinking about the incentives of incumbents to cater to the demands of interested parties,
we must consider what the local interest group and advocacy environment looks like from the perspective of the incumbent: is it, for example, dominated by a single large industry, such that the incumbent will have a strong incentive to cater to its state-independent policy preferences; or is it rich and heterogeneous, such that it more resembles the case of the neutral conduit. With respect to the latter, we have seen in the case of criminal justice policy how changes in the informational environment may dramatically change the incentives of incumbents.
Appendix: Some Related Studies

In this Appendix, we discuss some prior research that investigates the provision of information to voters by third parties in a TPA framework. Although not explicitly framed using the language of fire alarm oversight, several papers may be framed using our typology of neutral conduits, interested third parties, and challengers as sentinels. We exclude models in which challengers take actions that are revealing only about themselves but not about the incumbent ([Dewan and Hortalla-Vallve 2019]), models in which an outside player can engage in signal jamming preventing voters from learning about the incumbent or the state of the world ([Hirsch and Kastellec 2019]), and models with fire alarms but no electoral accountability ([Beim, Hirsch, and Kastellec 2014], [Prendergast 2003], see also [Gailmard and Patty 2017]). We divide the remaining papers into those in which sentinels are neutral conduits, interested parties, and challengers.

Neutral Conduits. The most obvious (potentially) neutral conduit is the media. The role of the media in electoral accountability is studied in several TPA papers, though often the question is the effect of media bias. An influential early paper is [Besley and Prat 2006]. The setting is pure adverse selection: some politicians are good and some are bad in a highly abstract sense. Voters can see neither actions nor consequences, so absent a media report they are in the NT information environment. The media probabilistically learn if the politician is the bad type and can make a hard information report to voters. If so, the voters move from NT to (effectively) the NA information environment, which will lead them to correctly fire the bad incumbent. However, much of the action in the model concerns bribes from the incumbent to the conduit, to silence a “bad type” report. Not surprisingly, government capture of the media is not in the voters’ interest; the paper goes well beyond this simple point to identify situations in which government capture of the media is more likely and less likely.

One of the most sophisticated investigations of media sentinels is [Wolton 2019]. This complex model focuses on preference alignment rather than skill. In the model, no politician type completely shares the preferences of the voter but some are worse for the voter than others. The voter begins in the NC environment (sees the incumbent’s action but does not see the consequences of the action). In one scenario, a media source acts as a neutral conduit, sending both a cheap-talk message about the incumbent’s type and a hard signal about the state of the world (so the voter
learns consequences). In the resulting equilibrium, both kinds of messages are informative to the voter, who is much better off from the presence of the neutral conduit.

**Interested Parties.** Biased media can be a type of interested party. One source of media bias is the profit motive, which creates incentives for sensationalism that are explored in an interesting paper on the political economy of libel laws ([Gratton 2015](#)). We pass over this rich paper to focus on more "political" approaches.

[Warren 2011](#) offers a model in which the incumbent has hidden ideological preferences, which are state contingent (see also [Maskin and Tirole 2004](#) and [Demirkaya 2019](#)). As a result, unless the voters know the state of the world, they cannot be sure whether the incumbent is ideologically congruent with them. In the model, a biased media sentinel may expend effort to check the incumbent’s signal about the state of the world. The sentinel faces two kinds of payoffs: first, a market reward from uncovering and revealing a mis-matched incumbent; second, a pro-incumbent bias. As a result of the latter, the media sentinel may cover up bad news about the incumbent, but it may also work hard to find good news. The net effects for the voter depend sensitively on the relative incentives for the biased sentinel.

We return to [Wolton 2019](#), which also considers scenarios in which the media sentinel has political preferences different from the voter. The paper distinguishes between two distinct configurations of biased sentinels: a “balanced” configuration, that is, with two sentinels one left-biased and one right-biased, and a “biased” configuration, with a single biased sentinel. Somewhat more illuminating language might have been “competitive biased-sentinels” and “monopoly biased-sentinel.” The balanced (competitive) configuration turns out to be favorable for the voter, but only when signals use hard information. The cheap-talk signals are not helpful to the voter. The logic of the result strongly resembles our analysis of challengers in Section 3, below. In essence, an adversarial setting without lying is a very favorable environment for information revelation (see also Dwattripont and Tirole 1999). The biased monopoly configuration is much less favorable for the voters, as the cheap-talk message involves some mis-representation and the hard-information signal sometimes involves cover-ups (strategic silence). Again, our analysis of interested parties in Section 3 has some strong similarities.

**Challengers.** Several papers offer models with signals about incumbents from electoral com-

---

17Our summary does not do justice to this complex model and its sometimes subtle results.
petitioners. For example, in [Gordon, Huber, and Landa 2007] the costly decision of a challenger to enter a race signals to voters that the incumbent is likely a poor performer. This is communication from the challenger to be sure, but not quite the fire-alarm messages highlighted by Arnold. A series of papers study the role of the opposition in parliamentary democracies, for example, [Dellis 2005], [Lemon 2005], [Demirkaya 2019], [Schultz 1998] and [Stasavage 2007]. The mechanisms at work in these papers vary. For example, some models treat the opposition and government as locked in an expertise contest in front of the voters during active policy making – each contestant tries to convince the voters that it is more expert than its foe. So, the competition takes the stylized form of the government saying “We are going to do X” and the opposition party saying “That’s a bad idea, it would be better to do Y.” When the consequences of X are revealed, the voters can award an electoral laurel to the apparently more astute party. A somewhat similar approach can also work with preference conformity.

In contrast to the mechanisms in these papers, the fire-alarm reports from an Arnold-style challenger provide information about otherwise hidden matters that have already taken place, i.e., the incumbent’s action, the consequences of his action, or the state of the world that prevailed when the incumbent chose his action. So the reports have the form “Hey voters, the incumbent did an unpopular thing,” “Look at this, the country is going to hell in a hand basket [due to the incumbent’s actions/inaction],” or “The incumbent’s clearly incompetent – he misunderstood the situation and totally screwed up.” Few models feature challengers making fire-alarm reports like this.

Two working papers by Scott Ashworth and Kenneth Shotts do feature reports like this, in the process offering two very different frameworks for studying challengers and Arnold-style fire-alarms. The first, [Ashworth and Shotts 2011], is a pure moral hazard setting with state-matching. In the model, absent a fire-alarm the voter is in the NT environment. With a fire-alarm, she moves to the NC environment – the voter can see how hard the incumbent worked (and working harder tends to lead to better consequences). Let’s cast the model into the gardener setting to make the story concrete. Here, it is costly and troublesome for the gardener to learn whether the garden is wet or dry (turning the sprinkler on or off takes no work at all). Nonetheless, the gardener would intrinsically like to state-match – for instance, he can pilfer a few strawberries and tomatoes from an ultimately flourishing garden. But aside from this motivation, he just doesn’t want to work hard,
absent the fear of firing. The challenger can spy on the gardener and see whether he is working hard or goofing off. The challenger can then show hard evidence of sloth (say, footage from his cell phone camera) to the employer. Given proof the gardener was goofing off, the employer can fire the current gardener and replace him with the spy. Unfortunately for the challenger, creeping around spying is also hard work, and will avail him naught if the gardener is actually working hard. Under the circumstance, can a good accountability equilibrium hang together – that is, one in which the gardener works, the challenger spies, the employer retains a diligent gardener and replaces a goof-off with the spy, and the garden does pretty well? Ashworth and Shotts find that the answer is "yes, but it isn’t easy." The problem is the crudeness of the retain/fire mechanism. It is, in effect, a single arrow that must hit two targets: motivating the incumbent to work hard at governing, and motivating the challenger to work hard at incumbent-monitoring and proof-gathering. Furthermore, and perhaps not surprisingly, reports from the spy must be hard information; otherwise, they are useless.

In [Ashworth and Shotts 2014], the same authors take a completely different approach to challengers and fire-alarms. This quite complex model features adverse selection with skillful and less skillful politicians, who are office-seekers. Both can produce hard information about the state of the world, and producing a report is linked to skill and hard work. Absent any reports, the voters are in the NC state since they see the incumbent’s policy choice but do not know the state of the world. With even one report, the voter moves to the FT state – they know the incumbent’s action and whether he state-matched. But voters still may be unsure about the types of both the incumbent and the challenger because an unskillful incumbent may luck into state-matching while even a skillful challenger may bomb at producing a hard-information report. The model is too complex to summarize neatly; indeed, in our experience even simplified versions lead rapidly to opacity. Nonetheless, a tractable adverse selection model seems intuitively appealing for some offices, such as the American presidency. There, voters are apt to learn something about the president’s "type" from four years of governing, and perhaps even more from a campaign.

We conclude our brief tour of TPA models with an interesting government-and-opposition paper, [Kishishita 2019]. In this model, the opposition may take an action (a filibuster) that allows the media – a neutral conduit – time to discover bad news about the government. The actions by the opposition and the media are complements, for the action by the opposition has no electoral
consequence unless it allows the media to act, and the media cannot report unless the opposition buys it time to investigate. Although the opposition in the model does not look like an Arnold-style challenger, the complementarity between the sentinel and the press is suggestive. Future research on challengers might consider the complementarity between the reports of a sentinel-challenger and the media: action by non-biased media has the twin effects of publicizing and verifying the challenger’s fire-alarm, hence, giving it effect. Conversely, biased media may amplify possibly bogus reports.
References


Society for Civil Engineers. 2017. ”2017 Infrastructure Report Card: Infrastructure Scores a D+,” https://www.infrastructurereportcard.org/


