

Whistleblowing and Compliance in the Judicial Hierarchy

Deborah Beim*

Department of Political Science
Yale University
deborah.beim@yale.edu

Alexander V. Hirsch
Department of Politics
Princeton University
avhirsch@princeton.edu

Jonathan P. Kastellec
Department of Politics
Princeton University
jkastell@princeton.edu

January 15, 2014

Abstract

One way that principals can overcome the problem of informational asymmetries in hierarchical organizations is to enable whistleblowing. We evaluate how whistleblowing influences compliance in the judicial hierarchy. We present a formal model in which a potential whistleblower may, at some cost, signal non-compliance by a lower court to a higher court. A key insight of the model is that whistleblowing is most informative when it is rare. While the presence of a whistleblower can increase compliance by lower courts, beyond a certain point blowing the whistle is counterproductive and actually reduces compliance. Moreover, a whistleblower who is a “perfect ally” of the higher court (in terms of preferences) blows the whistle too often. Our model shows an important connection between the frequency of whistleblowing and the effectiveness of whistleblowing as a threat to induce compliance in hierarchical organizations.

Forthcoming in the American Journal of Political Science

*We thank Tom Clark, Joshua Fischman, Justin Fox, Sandy Gordon, Lewis Kornhauser, James Rogers, and participants at Princeton University’s Political Methodology Colloquium, Emory University’s Conference on Institutions and Law-Making, and the LSE-NYU Conference on Political Science and Political Economy for helpful comments and suggestions.

1 Introduction

A pervasive problem in hierarchical organizations is how principals can oversee agents in the absence of all the information that agents typically possess (Bendor, Glazer and Hammond 2001). One way to overcome this problem is to promote the transmission of information by enabling *whistleblowing* or *fire alarms*. Both individuals within organizations and aggrieved third parties may act as whistleblowers by passing information to superiors. For example, McCubbins and Schwartz (1984) explore Congress' dilemma in deciding how best to oversee executive agencies who may choose to implement policies that differ from what members of Congress want. Given a choice between patrolling for non-compliance (a "police patrol") and waiting for third parties to notify Congress of non-compliance (a "fire alarm"), it is often more efficient to wait for individual citizens or interest groups to act as whistleblowers and sound an alarm.

The question of how best to oversee agents also arises in the federal judicial hierarchy. Appellate courts with discretionary dockets have the ability to choose which cases to review— with this power, higher courts can reverse decisions by lower courts that they disagree with. Because judicial decisions are the outlet for judicial power, the ability to oversee and reverse subordinates in the judicial hierarchy is a key mechanism for ensuring greater compliance and uniformity in the law (Kornhauser 1995). With this discretion, however, comes the difficulty of deciding which cases to select for review. Because American courts are tasked with judging only cases or controversies that come to them, it is difficult for them to conduct police patrols. Instead of seeking out disputes, courts must wait for cases to come to them.¹ Accordingly, higher courts must rely on "fire alarms" to select the few worthy cases to review. Litigant appeals serve as a form of fire alarms, given that many litigants will appeal a decision to a higher court if they believe that it is incorrect. In addition, external actors (i.e. those not directly involved in a legal dispute) also can sound fire alarms. These include judges on

¹While Supreme Court justices can send signals to lower courts, litigants, and interest groups about particular issues they would like to hear, the justices still must rely on other actors to bring them cases.

lower courts who write dissenting opinions targeted at higher courts, interest groups who file amici briefs asking a higher court to correct what they perceive as mistakes by a lower court, and a request by the solicitor general for the Supreme Court to hear a case.

It is well established that higher courts are more likely to review cases when any of these fire alarms have been sounded—in other words, that the Supreme Court relies on cues that suggest a case is worthy of review. The presence of a dissent on a three-judge panel of the U.S. Courts of Appeals is correlated with the probability that a circuit court will review a panel decision *en banc* (George 1999, Giles, Walker and Zorn 2006) and that the Supreme Court will review the case (Tanenhaus et al. 1963, Perry 1991, Caldeira, Wright and Zorn 1999). Cases in which an amicus brief is filed are more likely to be reviewed by the Supreme Court (Caldeira and Wright 1988). Finally, the Court is more likely to take cases when the solicitor general suggests review, since her selective involvement indicates that a case is important (Bailey, Kamoie and Maltzman 2005). But the mechanisms by which this information is useful to a higher court—and the incentives those mechanisms create for lower court judges—have been underexplored. If whistles increase the probability of higher court review, why don't lower court judges preempt them *ex ante* by accommodating potential whistleblowers? Similarly, why do some legal actors have more influence on the Supreme Court's case selection than others?

In this paper, we present a formal model of judicial decision making that evaluates how whistleblowing influences decision making in the judicial hierarchy. In our model, a lower court initially hears a case and decides whether to rule in a manner compliant with a higher court's known preferences. A *potential whistleblower*—which may represent a judge or a third party—then decides whether to send a costly public signal (i.e. to “blow the whistle”) in the form of a dissent, appeal or petition. The case then moves to a higher court, who observes the lower court's decision as well as this public signal if it is sent, but not the specific case facts. The higher court can then choose to review the case, learn the case facts, and potentially reverse the lower court's ruling. However, there is a cost to review, and

this cost is initially unknown to the other players. Importantly, we consider whistleblowers who are motivated by the desire to see case outcomes with which they disagree overturned. Moreover, because the higher court learns the facts upon review, the whistleblower never blows the whistle when the higher court would uphold the majority’s decision; thus, the whistleblower is always truthful. As a result, our model focuses on how to maximize the *informativeness* of these truthful signals.

The main results of the model are as follows. First, the presence of a potential whistleblower—and the associated threat of blowing the whistle—can increase compliance *ex ante*, by causing the lower court to vote against its preferred outcome more than it would in the absence of such a whistleblower. In such instances, the whistleblower has no need to blow the whistle.

Second, there is a *limit* on the extent to which the threat of whistleblowing can effectuate compliance by the lower court; it can never compel the lower court to fully comply in all cases. The reason is simply that the higher court’s review time is costly. Because of this, there are always cases where the lower court cares sufficiently about the outcome—and the higher court cares sufficiently little, compared to the cost of review—that the lower court is willing to risk review and reversal by ruling non-compliantly.

Third and most surprisingly, if the whistleblower chooses to blow the whistle on cases beyond this limit, then in equilibrium there is a “kickback” effect—compliance by the lower court is actually *reduced*. The reason for this effect follows from a fundamental property of whistleblowing—it is most effective for inducing the higher court to review a case when it is rare. We formally demonstrate how blowing the whistle more often reduces its informational value to the higher court, which in turn diminishes the likelihood that blowing the whistle will trigger review. The surprising equilibrium effect of this property is that blowing the whistle too much will eventually diminish the effectiveness of the *threat* of whistleblowing so much that the lower court will actually comply less than it would if the whistle was blown less frequently.

An important additional implication of this property is that compliance by the lower court

is maximized with *intermediate* whistleblowing—whistles that occur frequently enough that the lower court complies in many cases where they would not in the absence of a potential whistleblower, but not so often that blowing the whistle loses its informative value. When intermediate whistleblowing holds, blowing the whistle is used as a threat to constrain the lower court from engaging in more severe instances of non-compliance, but whistleblowing does not occur so often that the effectiveness of that threat is excessively diminished.

This “limit to compliance” generates a number of surprising implications and comparative statics. First, with respect to review, the signals of legal actors who are intrinsically more willing to blow the whistle—either because they are ideologically distant from the lower court or because blowing the whistle is less costly for them—will be less effective at inducing review. Second, with respect to compliance, the preferences of whistleblowers have a *non-monotonic* effect—their willingness to blow the whistle first increases compliance by the lower court, but eventually decreases compliance as more extreme whistleblowers become too willing to blow the whistle. Relatedly, the higher court does not benefit from having a “perfect ally” in the sense of having a whistleblower who perfectly shares its preferences. In fact, even a perfect ally whose costs and benefits are perfectly aligned with the higher court blows the whistle too often. The reason is that the decision to blow the whistle is driven by the immediate costs and benefits, and the whistleblower does not internalize the negative informational consequences of blowing the whistle too often.

In its focus on compliance, the model necessarily has a number of limitations. First, we focus on cases where the existing law is clear; thus, we ask how the *dispositions* issued by lower courts comply with the known preferences of a higher court, and focus on blowing the whistle solely as a signal about case facts. The incentives we study might look different in a model of *law creation*, where a higher court’s preferences are either unknown or unformed (see e.g. Baker and Mezzetti 2012, Carrubba and Clark 2012), and consequently lower courts have much more discretion in their decision making. Second, our model does not consider how the potential for dissent may influence bargaining on appellate courts over choices that

are not binary, such as the content of a legal rule (Cameron and Kornhauser 2010, Carrubba et al. 2012). In these situations, majorities and potential dissenting judges might compromise on moderate decisions.

Our model contributes to a growing literature on the incidence and effects of whistleblowing in institutions (Lupia and McCubbins 1994, Epstein and O’Halloran 1995, Hopenhayn and Lohmann 1996, Prendergast 2003, Ting 2008). In several of these models, however, the presence of a whistle is assumed; by contrast, we allow it to emerge endogenously. Moreover, while most of these studies have emphasized the *ex post* effects of whistleblowing—that is, how whistleblowers can inform superiors of possible non-compliance or mismanagement by an agent—our model clarifies the importance of the *ex ante* effect of whistleblowing. In this sense our paper complements the model of stovepiping presented in Gailmard and Patty (2013), which shows that the presence of a whistleblower can induce an agent to transmit information to the principal that she otherwise would not. Our evaluation of both the *ex ante* and *ex post* effects of whistleblowing also corresponds to the recent literature on laws designed to increase the oversight of cartels by granting leniency to members who blow the whistle on other members (Spagnolo 2008). Such laws have created both an *ex ante* cartel-deterrence effect and an *ex post* cartel-detection effect (Miller 2009). Finally, our result that intermediate whistleblowing is most effective in inducing compliance dovetails with the model in Takáts (2011), which shows that intermediate fines for banks that fail to report suspicious transactions are most effective in curtailing money laundering.

From the perspective of judicial politics, the notion of judicial whistleblowing was introduced by Cross and Tiller (1998), who found that judges on three-judge panels tended to vote differently depending on the preferences of their colleagues in a given case. Since then, scholars have extensively studied the phenomenon of “panel effects,” in which the propensity of a judge on a three-judge panel to vote liberally increases with every Democratic appointee she sits with, and vice versa (Revesz 1997, Sunstein et al. 2006). This literature has been largely empirical, although there have been a few theoretical exceptions (Kastellec 2007, Fis-

chman 2011, Spitzer and Talley 2013). Our model provides further theoretical foundations for understanding panel effects on the Courts of Appeals.

Our model also complements existing work on the judicial hierarchy more generally. Early models of the hierarchy evaluated courts on each level of the hierarchy as unitary actors (McNollgast 1995, Cameron, Segal and Songer 2000, Spitzer and Talley 2000). Lax (2003) considered a multimember Supreme Court to understand the effect of the “Rule of 4” on the Supreme Court’s interactions with the lower courts. The model in Kastellec (2007), upon which our model is built, maintained a unitary Supreme Court but introduced a two-player lower court in order to understand the relationship between panel effects and compliance. Whereas in Kastellec (2007), the threat of dissent was at times sufficient to induce full compliance by a lower court, the equilibrium behavior in the current model is more nuanced. Finally, our model shares some similarities with the theory presented in Daughety and Reinganum (2006). They do not focus on issues of compliance, but like our model they consider the role of dissent in providing information to a higher court.

2 The model

Players and Cases There are three players in the model: a higher court, H , a lower court L , and a potential whistleblower (henceforth simply “whistleblower”) W . L represents a majority bloc of judges, who we assume behave as a unitary actor. W represents an actor outside this majority who has the potential to issue a costly signal in the form of a dissent. W can represent a single judge on a three-judge panel, with L representing a two-judge majority. W can also represent an interest group filing an amicus brief or the solicitor general requesting the Court to hear the case. For ease of exposition, when there is a potential for pronoun confusion, we refer to H as “he,” L as “they,” and W as “she.”

The play of the game determines the outcome of a case, the facts of which map continuously onto a single dimension. The court makes either a “liberal” or “conservative” decision, denoted by *lib* and *con*, respectively. For example, Cameron, Segal and Songer (2000), Lax (2003), and Kastellec (2007) all describe the case space in terms of search-and-seizure cases.

In those models, the corresponding outcomes are a search is either held reasonable (the “conservative” outcome) or unreasonable (the “liberal” outcome).

The facts of the case map onto a unidimensional space X that determines the degree to which the liberal outcome is more appropriate; x denotes the case’s location on X . Facts that fall to the right of the space are more “liberal.” For example, in search-and-seizure law, the case space represents the degree of intrusiveness of a search, where cases that fall to the right are more intrusive.

Preferences and Utility The players care about case outcomes, and their preferences are described by an indifference point in the case space. With slight abuse of notation, we denote the players’ indifference points by L , W , and H . Each player prefers that all cases that map to the right of this indifference point receive the liberal outcome, and all to the left receive the conservative outcome. A player derives linear utility from the dispositional outcome: each receives $\frac{i-x}{2}$ from a ruling of *con* and $\frac{x-i}{2}$ from a ruling of *lib*, where i denotes each player’s respective indifference points. Thus, when $x < i$, the player prefers a ruling of *con*; when $x > i$ the player prefers *lib*. The loss from an incorrect decision is $|x - i|$. An indifference point can be thought of as a description of the player’s ideal legal outcome for every case.² To simplify the presentation, we assume throughout that L and H are located within $[0, 1]$, and in addition that the lower court is more liberal than the higher court, i.e. $L < H$. However, the results are symmetric: they continue to hold if one makes L more conservative than H and then transposes “liberal” for “conservative.”

Sequence of play Nature first randomly draws a set of case facts x distributed according to the CDF $F(x)$. $F(x)$ is assumed to be uniformly distributed on $[0, \bar{x}]$, where $\bar{x} \geq 1$. L and W then observe x , which is revealed to the lower court judges and interested parties as the case is presented in briefs and oral arguments. L decides whether to rule liberally or conservatively. After observing this choice, W decides whether to blow the whistle in the

²This conception of utility can comprise both legalistic and political goals, such as wanting to implement good law or pleasing external actors, like politicians who were responsible for a judge coming to the bench.

form of a public signal of cost $d \geq 0$. For ease of exposition, we refer to this signal as a “dissent” from this point forward; readers should bear in mind we mean dissent to encompass all types of fire alarms, including those from non-judges. Both L and W know the higher court’s preferences H , but do not know how much it would cost H to review L ’s decision.

The case then moves to H , who does not initially observe the case facts. However, he updates his prior beliefs based on the observable actions of L and W —specifically, L ’s disposition of the case, and whether W issued a dissent. This captures the informational asymmetry between a lower court that actually hears a case, and a higher court with a discretionary docket that initially has only limited information. H then decides whether to review and potentially reverse L ’s ruling. Since a review entails rehearing the case, we assume that H learns the case facts x upon review, and in addition pays a cost k . This cost captures both the time and resources a higher court must put into reviewing a case and the opportunity cost of hearing that case. Upon observing x the higher court then makes a final decision of whether to uphold or reverse L ’s ruling.

The cost of review k — H ’s cost of review—is assumed to be probabilistic and distributed according to a CDF $G(k)$, where $G(k)$ is a uniform distribution over $[0, \bar{k}]$ with $\bar{k} \geq 1$. This cost is known to H when he decides whether to review, but is initially unknown to both L and W . Substantively, L and W are uncertain about exactly how much H cares about getting the right disposition relative to the costs of hearing the case. Thus, they always entertain the possibility that H will choose not to review simply because his costs are too high. This uncertainty creates the possibility that L may rule non-compliantly despite knowing that she will trigger a dissent and raise the risk of review. We also allow \bar{k} to vary, thus allowing L and W to be more or less unsure of the cost to H of taking the case. Thus, paired with H ’s initial uncertainty over x , we assume a two-way informational asymmetry.

The cost of reversal Finally, we assume that costs accrue to L and W if L ’s initial ruling is reversed by H . If a reversal occurs, then L suffers a sanctioning cost $\varepsilon > 0$. This cost could capture, for example, the reputational penalty that a judge incurs when he is reversed.

When L is reversed, W is also assumed to suffer a cost $\alpha\varepsilon$. Informally, α captures the extent to which W 's fate is linked to L 's. On a three-judge panel where W represents a single judge, we argue that α is positive due to the fact that some costs of reversal fall on the entire court. Such costs include the cost of having to rehear a case on remand, and the court's general reputational cost. For third parties like litigants and interest groups, α could be low or even 0.³ Thus, to dissent W must be sufficiently opposed to the disposition to be willing to bear the costs of both dissenting and of being linked to L 's reversal.

3 Preliminary Analysis

In this section we begin characterizing Perfect Bayesian Equilibria of the model. In our analysis, we restrict attention to equilibria where a costly signal by the whistleblower increases the probability of review because it informs the higher court that relatively more severe non-compliance occurred. For clarity, in the remainder of the paper we refer to this class of equilibria as simply the “equilibria.”⁴

At the most general level, W creates a potential problem for L and a potential benefit for H (though there are certain scenarios, as we discuss, where the presence of W does not materially affect the play of the game). L must worry about whether W will dissent from a non-compliant decision. H , in turn, can use both the presence of a dissent, and the incentives created by the threat therein, to update his beliefs both about the likelihood that L is not complying and the severity of the possible non-compliance.

Before moving to the analysis, we foreshadow the general form of the equilibria. In theory, the form of each player's strategy could be complex; L must choose a ruling for every possible case fact, W must choose whether to dissent for every potential case fact and ruling, and

³We do not rule out the possibility that $\alpha < 0$, in which case the whistleblower benefits when L is reversed *irrespective* of her policy preferences. However, we argue that within most institutions $\alpha \geq 0$.

⁴As in most signaling models, there sometimes also exist equilibria where the meaning of the costly signal is reversed. In such equilibria, a “dissent” signals that non-compliance occurred but was minor and *reduces* the probability of review, while the lack of a dissent signals that the ruling was either severely non-compliant or fully compliant (with more weight on the former), and *raises* the probability of review. These “reversed” equilibria are not always present, and we omit consideration from the main text (see the Appendix for details). As we show in Proposition 2, equilibria of the class we consider always exist.

H must decide whether to review following every observable history of rulings and dissents. However, despite this potential complexity, all equilibria (in the class we consider) take the following simple and easily interpretable form. All proofs are gathered in the appendix.

Lemma 1 *All equilibria can be described by **cutpoints** $(c_L^*, c_W^*, \phi_{nd}^*, \phi_d^*)$, where $c_L^* \in [L, H]$ and $c_W^* < \min\{W, H\}$. Each of L and W 's actions in equilibrium depend on whether the case facts fall to the left or right of their respective cutpoints. H 's action—his choice of whether to review—depends on whether his cost of review falls above or below some threshold, the value of which is determined by whether or not W dissents. Specifically,*

1. *The lower court L rules liberally for $x \geq c_L^*$ and conservatively otherwise. When $x \in [c_L^*, H]$ this ruling is non-compliant.*
2. *The potential whistleblower W never dissents following a conservative ruling, and dissents following a liberal ruling when the facts are sufficiently conservative ($x \leq c_W^*$).*
3. *The higher court H never reviews conservative rulings, and sometimes reviews liberal rulings. Specifically, he reviews a liberal ruling i.f.f. $k \leq \phi_d^*$ when W dissents, and i.f.f. $k \leq \phi_{nd}^* < \phi_d^*$ when W does not dissent.*

The form of the equilibria is depicted in Figure 1; for reference, we label the four key regions, and refer to them below in the text. The structure of every equilibrium is summarized by four quantities $(c_L^*, c_W^*, \phi_{nd}^*, \phi_d^*)$. The lower court L rules conservatively when the facts are sufficiently conservative (Region 1, where $x \leq c_L^*$) and liberally otherwise (Regions 2-4, where $x \geq c_L^*$). When $x \in [L, c_L^*]$ (Region 1), their policy preferences would lead them to rule liberally but they choose to comply with the higher court's preferences and rule conservatively. When $x \in [c_L^*, H]$ (Regions 2 and 3) they choose to rule liberally, thus issuing a non-compliant decision.

In those non-compliant rulings where the facts are most conservative (Region 2, where $x \in [c_L^*, c_W^*]$), the whistleblower dissents. This informs the higher court that non-compliance has occurred and provides some information about its severity, thereby incentivizing him

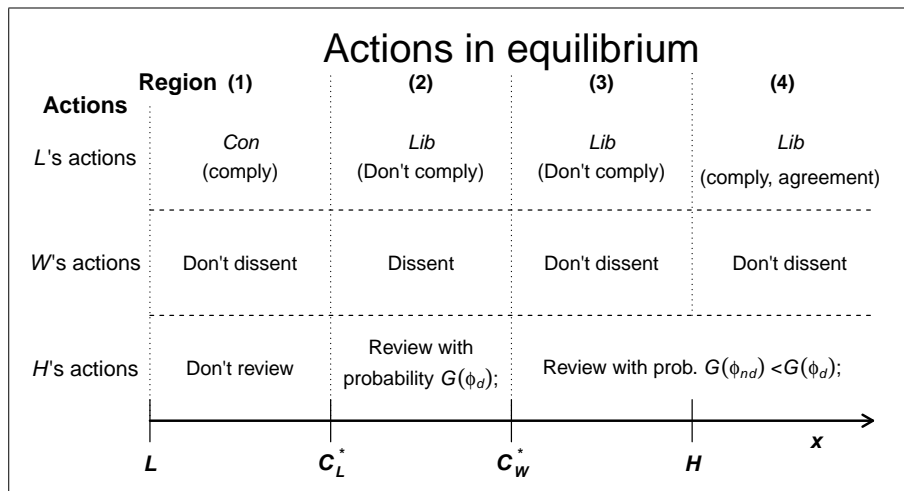


Figure 1: Summary of actions in equilibrium.

to review whenever the cost of review k is less than the threshold ϕ_d^* . Finally, for the remainder of rulings where the non-compliance is less severe (Region 3, i.e. $x \in [c_W^*, H]$), the whistleblower does not report it. This silence leaves H uncertain about whether the ruling involved less severe non-compliance (i.e. the case is in Region 3), or was actually compliant (i.e. the case is in Region 4). He thus reviews only when the cost k is below some lower threshold $\phi_{nd}^* < \phi_d^*$.

Before proceeding with the analysis, we emphasize that a player's strategy should not be interpreted as "choosing a cutpoint." Rather, at the moment that each player makes their decision, they choose an optimal action given the observable history and their expectations about the other players' actions. The fact that these actions can be described by cutpoints is a *consequence* of equilibrium, not an assumption about the strategy space. For example, the whistleblower decides whether to dissent after observing the realization of the case facts x and L 's ruling of *lib* or *con*. She does so based on her fixed expectations about the probability that dissenting will trigger a review, derived from the higher court's strategy. These incentives imply that her best responses can be described by a cutpoint c_W separating the liberal rulings on which she would dissent from those on which she would not.

To characterize how these equilibrium cutpoints are jointly determined, we proceed as follows. In the remainder of this section, we characterize each player's best responses as a function of the others' strategies; this provides intuition for the basic incentives under-

lying equilibrium. In the subsequent section, we characterize equilibria, discuss important properties, and present our main results.

3.1 The Value of Review to the Higher Court

We begin by examining the higher court’s review decision conditional on L and W ’s strategies and actions. H assesses the value of review by observing the ruling and whether W dissented, and then updates his beliefs about the case facts based on L and W ’s strategies. In equilibrium his review cutpoints ϕ_d^* and ϕ_{nd}^* must equal these assessments.

A Conservative Ruling For strategy profiles of the form in Lemma 1, H ’s assessment of the benefit of reviewing a conservative ruling is straightforward to compute: there is none. Since L is more liberal than H , any case on which L is willing to rule conservatively ($x < c_L$) is also one for which H would prefer a conservative ruling ($x < H$). This is the “Nixon goes to China” result established in Cameron, Segal and Songer (2000)—if a more extreme lower court reaches a conclusion that goes against the direction of their (relative) ideological bias, the higher court can be sure he would rule the same. Thus, H ’s best response is to not review any conservative rulings. Recall that the results are symmetric: a more liberal H responds identically to liberal decisions by a more conservative L .

A Liberal Ruling Accompanied by a Dissent While H always agrees with L ’s conservative decisions, he may not always agree with L ’s liberal decisions. After liberal decisions, H ’s inferences are as follows. Consider first when L makes a liberal ruling that is accompanied by a dissent from W . In this circumstance, H can infer that the case facts are in $[c_L, c_W]$.⁵ Intuitively, H knows that non-compliance occurred and that it was relatively severe, but does not know the precise severity. Were he to review the case and learn the facts, he expects he would reverse the decision for sure and gain utility $H - x$ (recall that the utility from an incorrect decision is $-\frac{H-x}{2}$ and from a correct one is $\frac{H-x}{2}$). His *expected utility gain*

⁵When $c_W < c_L$ dissent is off-path and we must specify beliefs. To preserve continuity we assume that H believes W to have deviated from her strategy by dissenting when the case facts were precisely at c_L .

from review is therefore,

$$\phi_d(c_L, c_W; H, F(\cdot)) = H - E[x | x \in [c_L, c_W]], \quad (1)$$

or the difference between his indifference point and the expected case facts, conditional on x being in $[c_L, c_W]$.⁶ A best response by H thus requires that he review a liberal ruling accompanied by a dissent if and only if his realized cost k is less than $\phi_d(c_L, c_W; H, F(\cdot))$.

A Liberal Ruling Without a Dissent When L rules liberally but W does not dissent, H must consider the possibilities both that less severe non-compliance occurred ($x \in [c_W, H]$), or that the case facts were sufficiently liberal that the ruling was actually compliant ($x > H$). If L indeed failed to comply, then as before H 's gain to review is $H - x$, since he would discover the non-compliance upon review and reverse the ruling. However, if the ruling was actually compliant, then upon review H would make no change to the disposition, and the gain to review would be 0 but he would have paid the cost k of reviewing. His *expected utility gain* from review is therefore:

$$\phi_{nd}(c_L, c_W; H, F(\cdot)) = P(x \in [c_W, H] | x > c_W) \cdot (H - E[x | x \in [c_W, H]]) \quad (2)$$

As before, a best response by H requires that he review a liberal ruling unaccompanied by a dissent *i.f.f.* his cost k is less than $\phi_{nd}(c_L, c_W, H, F(\cdot))$.⁷

The Effectiveness of Dissent The effect of dissent on the higher court's beliefs is both to raise the probability that non-compliance occurred from $P(x \in [c_W, H] | x > c_W)$ to 1, and to shift the possible case facts to the left from the interval $[c_W, H]$ to the interval $[c_L, c_W]$. Together these updates *increase* the court's expected gain to review from $\phi_{nd}(\cdot)$ to $\phi_d(\cdot)$. However, because the higher court's opportunity cost of review is *probabilistic* and unknown to L and W , this increase does not generate review with certainty. Instead, it increases the likelihood of review from $G(\phi_{nd}(c_L, c_W; \cdot))$ to $G(\phi_d(c_L, c_W; \cdot))$; in particular, when $\phi_{nd}(c_L, c_W; \cdot) < k < \phi_d(c_L, c_W; \cdot)$, H will *only* review if there is a dissent. Moreover,

⁶Given our assumption about off-path beliefs, the general expression to account for $c_W < c_L$ requires simply substituting $\max\{c_W, c_L\}$ for c_W .

⁷Again to account for the case of $c_W < c_L$ substitute $\max\{c_W, c_L\}$ for c_W .

although dissent always increases the probability of review, the exact increase depends on the *frequency of dissent* (i.e. the location of c_W) via its influence on H 's beliefs about the expected gain from review.

An important feature of the model is that dissenting on a larger set of cases (i.e. a higher c_W) *reduces* the likelihood of review following a dissent. The reason is this: although W dissents only when non-compliance has actually occurred, she could be dissenting on a wider set of cases, including some with relatively mild non-compliance, or only on a narrow set of cases with relatively severe compliance. When W dissents more, the higher court's expected gain from review following dissent $\phi_d(c_L, c_W; \cdot) = H - E[x | x \in [c_L, c_W]]$ is lower because the additional cases on which W is dissenting are those for which the non-compliance is least severe. This translates into a lower equilibrium probability of review following dissent. We summarize these results in the following lemma:

Lemma 2 *In a best response, the higher court:*

- *never reviews a conservative ruling;*
- *reviews a liberal ruling with a dissent i.f.f. $k < \phi_d(c_L, c_W; \cdot)$;*
- *reviews a liberal ruling without a dissent i.f.f. $k < \phi_{nd}(c_L, c_W; \cdot) < \phi_d(c_L, c_W; \cdot)$.*

Moreover, the higher court's probability of review following a dissent $G(\phi_d(c_L, c_W; \cdot))$ is strictly decreasing in the whistleblower's dissent cutpoint c_W .

3.2 The Benefits of Dissent to the Whistleblower

We now consider the incentives of the whistleblower. In our model, the whistleblower dissents to persuade the higher court to review and reverse a non-compliant liberal ruling. Consequently, she will never dissent from a liberal ruling on cases $x \geq H$ where the higher court also prefers the liberal disposition, since he will never reverse them. On cases where the higher court would reverse a liberal ruling upon review ($x < H$), the whistleblower would realize a net gain of $(W - x) - \alpha \cdot \varepsilon$ if she succeeded in inducing a review; this is the utility gain from reversing a non-compliance liberal disposition, less W 's share of the sanctioning cost that falls on the lower court upon reversal. W 's dissent will succeed in inducing a review

that would otherwise have not occurred when $k \in [\phi_{nd}, \phi_d]$; that is, when the higher court's cost of review falls between his threshold for reviewing a liberal ruling absent a dissent and with a dissent. Thus, W will find it worthwhile to dissent when

$$((W - x) - \alpha\varepsilon) \cdot (G(\phi_d) - G(\phi_{nd})) > d, \quad (3)$$

which generates the following best response behavior:

Lemma 3 *When $\phi_d > \phi_{nd}$, the whistleblower's best response is to dissent if and only if $x < \min\{c_W(\phi_d, \phi_{nd}; W, d, \alpha), H\}$, where*

$$c_W(\phi_d, \phi_{nd}; W, d, \alpha) = (W - \alpha\varepsilon) - \frac{d}{G(\phi_d) - G(\phi_{nd})} \quad (4)$$

The whistleblower's dissent behavior in a best response therefore takes the form of a cutpoint $\min\{c_W(\phi_d, \phi_{nd}; W, d, \alpha), H\}$ that has intuitive properties. First, it increases (i.e., leads to more dissents) both as W becomes more conservative and as the probability that dissent is *pivotal* $G(\phi_d) - G(\phi_{nd})$ increases. Second, it decreases (i.e., leads to fewer dissents) both as W 's share of the reversal sanction α rises, and as the cost of dissent d grows.

3.3 The Risk of Review to the Lower Court

Finally, we examine to the calculus of the lower court. For L , the benefit of non-compliance on a case $x \in [L, H]$ is obtaining the liberal outcome. The cost is the risk of review, which reverses this outcome and generates a sanction of cost ε . A key component of its decision is therefore the likelihood of review. If this were some fixed probability q , then the gain from non-compliance would be $(1 - q) \cdot (x - L)$ (the probability of no review times the gain from a liberal ruling), and the cost would be $q \cdot \varepsilon$ (the probability of review times the sanctioning cost). In a best response, L would rule non-compliantly whenever x is greater than

$$x^*(q; L, \varepsilon) = L + \varepsilon \left(\frac{q}{1 - q} \right). \quad (5)$$

This function is increasing in q ; the lower court would comply more with the preferences of the higher court when the probability of review is higher.

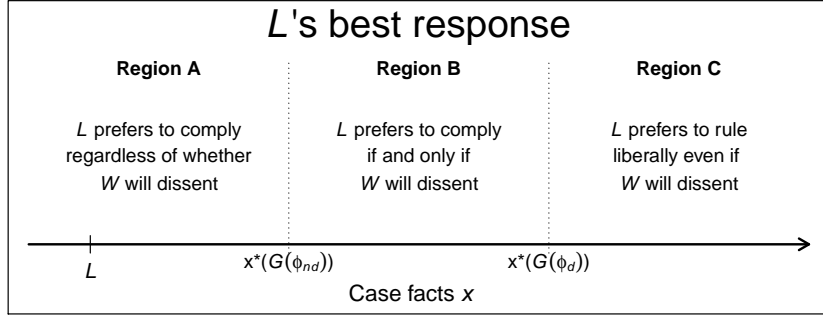


Figure 2: L 's best response.

However, the probability of review of a liberal disposition is not fixed; it is determined by the whistleblower's dissent behavior. Specifically, the probability of review is $G(\phi_d)$ on those cases where the whistleblower will dissent ($x < c_W$), and $G(\phi_{nd}) < G(\phi_d)$ on those cases where the whistleblower will not dissent ($x > c_W$). Hence, from the perspective of the lower court the case space can be divided into three regions, as shown in Figure 2:

- Region A: cases $x < x^*(G(\phi_{nd}); \cdot)$, on which the facts are sufficiently conservative that L prefers to comply *regardless of* whether W will dissent;
- Region B: cases $x \in [x^*(G(\phi_{nd}); \cdot), x^*(G(\phi_d); \cdot)]$ on which the case facts are intermediate and L prefers to comply *only if* they expect non-compliance to trigger a dissent.
- Region C: cases $x > x^*(G(\phi_d); \cdot)$ on which the case facts are sufficiently liberal that L prefers to rule liberally *even if* W will dissent.

The middle region (Region B) comprises the set of cases on which the whistleblower's *threat* of dissent can induce compliance by L . Outside of this region, the lower court prefers to either comply (Region A) or risk reversal (Region C) regardless of whether the whistleblower will dissent. Consequently, the lower court's best-response behavior is as follows.

Lemma 4 *In a best response, the lower court L uses cutpoint,*⁸

$$c_L(c_W, \phi_d, \phi_{nd}; L, \varepsilon) = \begin{cases} x^*(G(\phi_{nd}); \cdot) & \text{if } c_W < x^*(G(\phi_{nd}); \cdot) \\ c_W & \text{if } c_W \in [x^*(G(\phi_{nd}); \cdot), x^*(G(\phi_d); \cdot)] \\ x^*(G(\phi_d); \cdot) & \text{if } c_W > x^*(G(\phi_d); \cdot) \end{cases} \quad (6)$$

⁸Equation 6 may be written more succinctly as $c_L(c_W, \phi_d, \phi_{nd}) = \min\{H, x^*(G(\phi_d)), \max\{x^*(G(\phi_{nd}), c_W)\}\}$. Note that there is also an upper bound H , since L will never rule conservatively on a case where H prefers the liberal ruling.

4 Properties of Equilibria

We now characterize equilibria of the model and present our main results. Formally, an equilibrium requires that the higher court, whistleblower, and lower court be jointly best-responding to each other's strategies. That is,

Lemma 5 *Cutpoints $(c_L^*, c_W^*, \phi_{nd}^*, \phi_d^*)$ are an equilibrium i.f.f. they satisfy Lemmas 2 - 4.*

In the remainder of this section we provide a more precise characterization of the equilibria by solving for cutpoints that jointly satisfy the necessary and sufficient conditions. In doing so, we describe equilibrium patterns of behavior and derive comparative statics. To identify these cutpoints we proceed in two steps. First, we fix the whistleblower's cutpoint c_W , and characterize how that fixed level of dissent affects the equilibrium incentives of the lower and higher courts. Formally, we characterize the unique *partial equilibrium* level of compliance by the lower court $c_L^*(c_W)$ when it and the higher court are jointly best-responding to each other and to c_W . Second, we use this characterization to solve for equilibrium values of the whistleblower's dissent cutpoint c_W^* .

4.1 The Limits to Dissent

To characterize how the whistleblower's cutpoint c_W affects the lower court's equilibrium level of compliance $c_L^*(c_W)$, we employ two benchmark results. Both benchmarks consider a two-player game played only between the lower and higher courts. In the first, the higher court has *no information* about the case facts (as in the main model). In the second, the higher court has *complete information* about the case facts. Since the function of the whistleblower is to provide the higher court with information about the case facts, these benchmarks allow us to understand her contribution to the informational and strategic environment.

In the *no information* benchmark, H reviews solely on the basis of his beliefs about how much the lower court is complying with his preferences. In this game (as in the main model), L complies if and only if the case facts fall below some equilibrium cutpoint, which we denote \underline{c}_L . This cutpoint is defined by the equality $\underline{c}_L = x^*(G(\phi_{nd}(\underline{c}_L, \underline{c}_L; \cdot)); \cdot)$, because the higher

court's beliefs about the benefits of the review absent the whistleblower are equal to what her beliefs $\phi_{nd}(c_L, c_L)$ would be in the presence of a whistleblower who never dissents.

In the *complete information* benchmark, the higher court already knows the case facts and does not need to review to learn them; thus, the *only* purpose of review is to provide the opportunity to reverse a ruling known to be non-compliant. Again, L 's decision to comply is based on an equilibrium cutpoint, which we denote \bar{c}_L . This cutpoint is defined by the equality $\bar{c}_L = x^*(G(\phi_d(\bar{c}_L, \bar{c}_L; \cdot)); \cdot) = x^*(G(H - \bar{c}_L))$, because when the whistleblower uses cutpoint $c_W = c_L$ a dissent perfectly reveals that the case facts are exactly at $x = c_L$.

We can now use these two benchmarks to identify the whistleblower's contribution to the strategic environment in the following proposition:

Proposition 1 *The lower and higher court's joint best response behavior as a function of the whistleblower's dissent cutpoint c_W is characterized by three nonempty regions defined by the cutpoints $\underline{c}_L < \bar{c}_L$, which are both strictly interior to (L, H) .*

- **Region I** ($c_W < \underline{c}_L$): *The lower court is unaffected by the threat of dissent and sets compliance at $c_L^*(c_W) = \underline{c}_L$. The degree of compliance is **constant**, the probability of dissent is **zero**, and the probability of review following a dissent is **constant**.*
- **Region II** ($c_W \in [\underline{c}_L, \bar{c}_L]$): *The lower court complies just enough to avoid dissent by setting $c_L^*(c_W) = c_W$. The degree of compliance is **increasing**, the probability of dissent is **zero**, and the probability of review following a dissent is **decreasing**.*
- **Region III** ($c_W > \bar{c}_L$): *The lower court is partially affected by the threat of dissent and solves $c_L = x^*(G(\phi_d(c_L, c_W; \cdot)))$. The degree of compliance is **decreasing**, the probability of dissent is **increasing**, and the probability of review following a dissent is **decreasing**.⁹*

The three panels in Figure 3 depict the effect of the whistleblower's dissent cutpoint c_W on equilibrium levels of compliance, dissent, and review behavior. Specifically, Figure 3A

⁹This proposition does not require either $F(\cdot)$ or $G(\cdot)$ to be uniform.

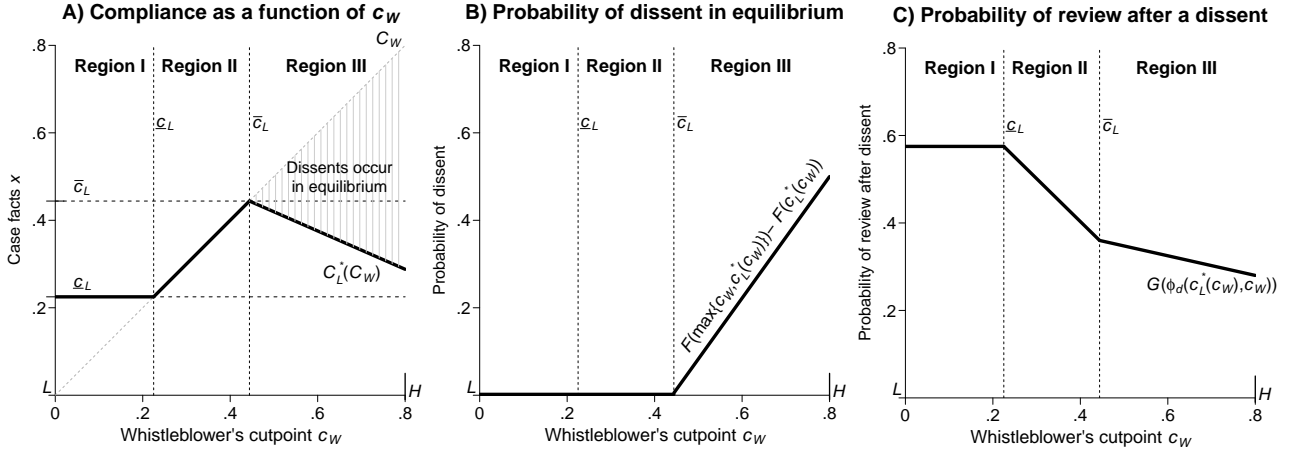


Figure 3: The effect of the whistleblower’s dissent cutpoint c_W on: **A)** compliance; **B)** dissenting; and **C)** review behavior. In each panel, Regions I and II are separated by \underline{c}_L , which is the cutpoint that the lower court would use if the higher court had no information about the case facts (in a game played just between L and H), while Regions II and III are separated by \bar{c}_L , which is the cutpoint the lower court would use if the higher court were fully informed about the case facts.

shows its effect on L ’s equilibrium compliance $c_L^*(c_W)$; Figure 3B shows its effect on W ’s probability of dissent $F(\max\{c_W, c_L^*(c_W)\}) - F(c_L^*(c_W))$; Figure 3C depicts its effect on H ’s probability of review after dissent $G(\phi_d(c_L^*(c_W), c_W; \cdot))$.

There are three key regions, separated by the cutpoints \underline{c}_L and \bar{c}_L from the two benchmarks. Beginning with L ’s compliance behavior, in the leftmost region (Region I, $c_W < \underline{c}_L$) there are *no whistleblower effects* on compliance by the lower court—it complies to the same degree it would absent the whistleblower in the no information benchmark. The reason is that all cases on which the whistleblower would dissent ($x < c_W$) are also ones where the lower court would comply absent the threat of dissent ($x < \underline{c}_L$). Consequently, in this region the probability of dissent is zero and the probability of review after a dissent is constant.¹⁰

In sharp contrast, the middle region (Region II, $c_W \in [\underline{c}_L, \bar{c}_L]$) exhibits *full whistleblower effects*. Here, the lower court avoids dissent by complying exactly on those cases ($x < c_W$) on which the whistleblower *would* dissent. Thus, L ’s compliance is increasing in the whistleblower’s cutpoint, while the probability of review after a dissent is decreasing. The probability of dissent remains zero, and no actual dissents are observed in equilibrium.

¹⁰Recall that H ’s off-path beliefs when a dissent is observed off-path is $x = c_L$.

The rightmost region (Region III, $c_W > \bar{c}_L$) exhibits *partial and diminishing whistleblower effects*. Here, the threat of dissent still induces the lower court to comply more than it would absent the threat ($c_L^*(c_W) > \bar{c}_L$)—this can be seen by comparing compliance in Regions I and III. However, on some cases for which L values the liberal ruling sufficiently highly ($x \in [c_L^*(c_W), c_W]$), it chooses not to comply *even knowing* that a dissent will be triggered. Thus, dissents are observed in equilibrium, and the probability of a dissent is increasing in the whistleblower’s cutpoint. More interestingly, within this region more dissenting by W (i.e. a greater c_W) has the counterproductive effect of *diminishing* W ’s influence on the lower court and generating even less compliance (i.e. a lower $c_L^*(c_W)$). This demonstrates that there is a “limit to dissent” as a tool for effectuating compliance by a lower court.

The preceding observations have surprising implications for how a whistleblower’s behavior affects compliance by a lower court.

Corollary 1 *When the lower and higher courts are jointly best responding, the whistleblowing cutpoint c_W that maximizes the lower court’s compliance is equal to $\bar{c}_L \in (L, H)$.*

Why is reporting of non-compliance on cases beyond \bar{c}_L not only ineffective for inducing additional compliance, but increasingly counterproductive? Because there is a chance that the cost of review will be too high for the higher court to be willing to review the case, there is an upper bound on how much the lower court can ever be induced to comply—this upper bound is \bar{c}_L , or the cutpoint from the complete information benchmark. For cases to the right of this cutpoint, the lower court would risk escaping reversal by the higher court *even if H had complete information about the case facts*, because of the possibility that the cost of review would be too high to make review worthwhile. When W dissents on cases to the right of this cutpoint (that is, when $c_W > \bar{c}_L$) L simply cannot be pushed to comply more. As a result, the *only* effect of this additional dissenting is to diminish the *informational value* of dissent by lumping more severe instances of non-compliance together with less severe instances. As characterized in Section 4.2, this lowers H ’s equilibrium beliefs about the expected severity of non-compliance upon observing a dissent (i.e. a lower $\phi_d(\cdot)$),

which causes H to respond to dissents less (i.e. a lower $G(\phi_d(\cdot))$) and ultimately results in *less compliance* by the lower court in equilibrium (i.e. a lower $c_L^*(c_W)$).

The fact that frequent dissenters are less effective is well understood by members of the judiciary—for example, Justice Ruth Bader Ginsburg has warned of the “danger of crying wolf too often” (Ginsburg 1990, 142), and Justice Harlan Stone wrote to Karl Lewellyn, “If I should write in every case where I do not agree with some of the views expressed in the opinions, you and all my other friends would stop reading them” (Murphy 1964, 62). Our model, however, demonstrates a previously unidentified implication of this effect; excessively frequent dissenting can actually *reduce compliance* by a lower court. The reason is that whistleblowers have both an *ex ante* and *ex post* effect on judicial outcomes. *Ex post*, their dissents increase the probability of review. *Ex ante*, the *threat* of their dissents increases the lower court’s incentive to comply. If a whistleblower excessively “squander[s] the ammunition of dissent,” (Llewellyn, Gewirtz and Ansaldi 1988, 1000, fn.10) she can dilute the effectiveness of that ammunition so much that a lower court is *more willing* to rule non-compliantly and risk review. Formally, increasing the whistleblowing cutpoint *above* \bar{c}_L (i.e. to the right in our model) leads to a concomitant shift in c_L^* towards *less compliance* (i.e. to the left). Substantively, when “friends” (that is, judges with similar preferences) stop reading dissents, “enemies” will be able to ignore whistleblowers’ threats.

4.2 Equilibrium Whistleblowing

Having characterized how the whistleblower’s dissent behavior affects the lower court’s equilibrium level of compliance $c_L^*(c_W)$, we now complete the analysis. Specifically, we first characterize equilibrium values of the whistleblower’s dissent cutpoint c_W^* , then describe when the whistleblower’s presence will effect the lower court’s equilibrium compliance behavior, and finally derive comparative statics in the whistleblower’s parameters.

Proposition 2 *Equilibria always exists and satisfy the following.*

1. A pair of cutpoints (c_L^*, c_W^*) are an equilibrium i.f.f. $c_L^* = c_L^*(c_W^*)$ and

$$c_W^* = \min \left\{ (W - \alpha\varepsilon) - \frac{d}{G(\phi_d(\max\{c_L^*, c_W^*\}, c_W^*)) - G(\phi_{nd}(\max\{c_L^*, c_W^*\}, c_W^*))}, H \right\}.$$

2. There exists an equilibrium that exhibits whistleblower effects ($c_L^* > \underline{c}_L$) i.f.f.

$$W > \underline{c}_L + \left(\alpha\varepsilon + \frac{d}{G(\phi_d(\underline{c}_L, \underline{c}_L)) - G(\phi_{nd}(\underline{c}_L, \underline{c}_L))} \right),$$

where \underline{c}_L is as characterized in Proposition 1.

The first part of the Proposition provides necessary and sufficient conditions for compliance and dissent cutpoints (c_L^*, c_W^*) to be an equilibrium; the lower court must be best-responding with $c_L^*(c_W^*)$, and the whistleblower must be best responding to this level of compliance. The second part of the Proposition states necessary and sufficient conditions for *whistleblower effects* to occur in equilibrium—that is, for the threat of the whistleblower’s dissent to increase lower court compliance relative to what would occur absent the whistleblower. The condition is that W be more conservative than the threshold that is defined in the proposition. The required threshold is increasing in both the cost of dissent d and the whistleblower’s share of the reversal sanction α .¹¹

Comparative Statics Proposition 2 does not rule out multiple equilibria, which complicates deriving comparative statics. Intuitively, the reason is that dissenting beyond the limit to compliance \bar{c}_L can be self-reinforcing; by decreasing the lower court’s compliance, such dissents also increase the higher court’s responsiveness to dissent, and consequently the whistleblower’s willingness to dissent. However, because our emphasis is on how much compliance can be achieved, for the remaining analysis we select the equilibrium in which the *maximum* level of compliance is sustained; we denote this quantity \tilde{c}_L .¹² This selection reflects a broader normative interest in institutional arrangements that align the rulings of lower courts with the preferences of higher courts.

We now describe how maximum equilibrium compliance changes as a function of the parameters directly affecting the whistleblower’s willingness to dissent.

Lemma 6 *Maximum equilibrium compliance \tilde{c}_L exhibits three consecutive regions as a function of the three parameters that directly affect the whistleblower’s willingness to dissent:*

¹¹This is derived by considering whether W would wish to dissent on any non-compliant cases when L uses cutpoint \underline{c}_L and H believes dissent to be maximally informative ($c_W = \underline{c}_L$).

¹²Formally, $\tilde{c}_L = \max \{c_L \text{ s.t. } \exists c_W \text{ s.t. } (c_L, c_W) \text{ are an equilibrium}\}$.

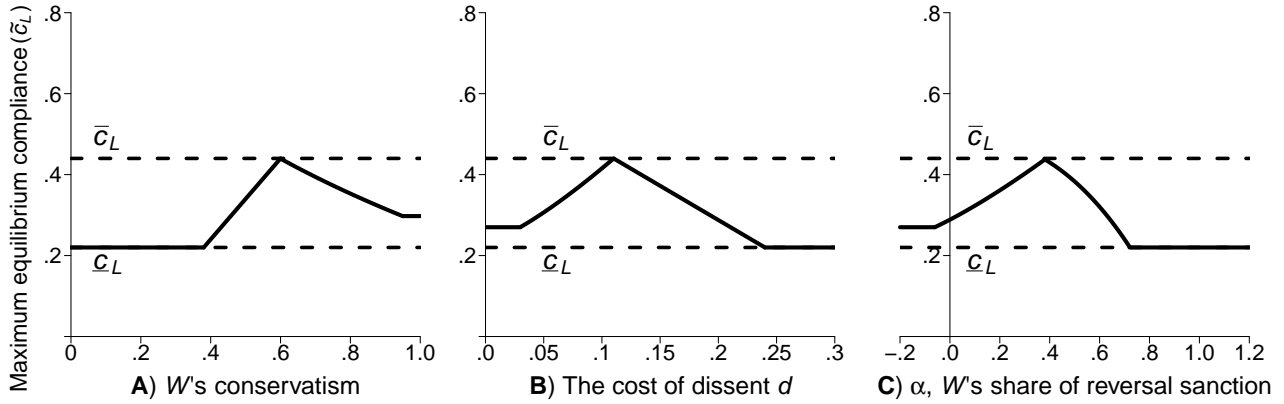


Figure 4: The maximum level of compliance, as a function of **A)** the whistleblower’s conservatism, **B)** the cost of dissent, and **C)** W ’s share of the reversal sanction. We assume $L = 0$, $H = .8$, $\varepsilon = .8$, $\bar{k} = 1$, $W = .9$, $d = .04$, and $\alpha = 0$, except we allow the latter three to vary in their respective plots. The dashed horizontal lines depict \underline{c}_L and \bar{c}_L .

her conservatism W , the cost of dissent d , and her share of the reversal sanction α . These regions mirror the regions associated with $c_L^*(c_W)$ in Proposition 1.

Thus, the effects of W , d , and α on maximum equilibrium compliance \bar{c}_L are essentially identical to the effect of the whistleblower’s cutpoint c_W on partial equilibrium compliance $c_L^*(c_W)$. Both sets of comparative statics exhibit three regions: one with no whistleblower effects, followed by a region with full whistleblower effects, followed by a region with partial and diminishing whistleblower effects. These results are illustrated in Figure 4, which is based on a numerical example with $L = 0$, $H = .8$, $\varepsilon = .8$, $\bar{k} = 1$, $W = .9$, $d = .04$, and $\alpha = 0$. Figure 4A depicts how the maximum equilibrium compliance varies as W becomes more conservative. Initially, W has no effect on L ’s cutpoint, and thus there are no whistleblower effects (i.e. $\bar{c}_L = \underline{c}_L$). Then a more conservative whistleblower increases compliance until it reaches the compliance limit (\bar{c}_L), at which point increasing W lowers \bar{c}_L . Figures 4B and 4C depict a similar effect of increasing either the cost of dissent (d) or the whistleblower’s share of the reversal sanction (α). Note that the order of the regions is reversed for the share of the sanction and the cost of dissent, since increasing either *decreases* the whistleblower’s willingness to dissent.

4.3 Optimal Whistleblowing and the “Ally Principle”

To conclude our analysis, we consider the ideal whistleblower from the perspective of the higher court. An influential organizing idea in the study of principal-agent relationships is

the *ally principle*, which states that “a principal is made best off by appointing as his or her agent the individual whose preferences over outcomes are most similar to those of the principal” (Gailmard and Patty 2012, 367). In our model, the ally principle clearly applies to the relationship between the higher court and the lower court—an L who is identical to H would always comply. However, we show in this subsection that it does *not* generally hold with respect to the relationship between the higher court and whistleblowers.

To understand what type of whistleblower the higher court would want in terms of *preferences* (i.e. W), we must first examine the higher court’s preferences over the whistleblower’s cutpoint (i.e. c_W). The higher court’s preferences over the latter are determined by two factors—his desire for accurate information about the severity of non-compliance (to minimize wasteful reviews), and his desire for the threat of dissent to induce compliance. However, Proposition 1 implies that these two considerations are at odds: when the whistleblower’s cutpoint is above the limit to compliance (i.e. $c_W > \bar{c}_L$), better information about the severity of non-compliance comes at the cost of further reducing the effectiveness of dissent for inducing compliance. This tradeoff generates the following result about the higher court’s preferences over the whistleblower’s dissent cutpoint.

Lemma 7 *When the lower and higher courts are jointly best responding, the whistleblowing cutpoint c_W that maximizes the higher court’s expected utility is strictly less than H and weakly greater than \bar{c}_L .*

The higher court thus *never* prefers that the whistleblower fully report all instances of non-compliance. For many parameter values (including our numerical examples), the negative effect of whistleblowing beyond \bar{c}_L on the lower court’s compliance is so severe that the higher court actually prefers the whistleblower to report non-compliance only up to \bar{c}_L .

What type of whistleblower would engage in such optimal intermediate whistleblowing? In general it is *not* a perfect ally of the higher court, in the sense of preferences. The reason is this: unlike in many settings where costs are shared between principals and agents, in the judicial setting whistleblowers alone pay the costs of issuing dissents, may suffer some of the

reversal costs that fall on the lower court, and are spared the higher court’s costs of wasteful reviews. For example, a whistleblower who shares H ’s preferences but for whom dissent is cheap will generally dissent too much because she does not suffer H ’s costs of review.

More surprisingly, however, the ally principle fails in our model *even when* the whistleblower and higher court’s costs and benefits are perfectly aligned.

Lemma 8 *When dissent and reversal are costless for the whistleblower, and the whistleblower internalizes the higher court’s cost of review, the higher court’s expected utility is maximized by a whistleblower whose preferences are strictly less than H .*

The failure of the ally principle under these conditions stems from the nature of dissent itself: it can only occur *after* the lower court has made its decision. Consequently, the whistleblower is unable to internalize the *equilibrium* consequences of her dissent behavior on the lower court’s compliance. Instead, she simply dissents based on the immediate costs and benefits, and will therefore dissent too much when her payoffs are perfectly aligned with the higher court.

As described above, judges themselves do indicate a concern with the broader consequences of dissenting too much. However, it is unclear whether they can mitigate their incentives to do so by “tying their own hands.” Although outside the scope of our model, the repeated interactions that occur between higher courts and potential whistleblowers may provide one solution to this problem: judges may be able to develop a *reputation* for only whistleblowing in egregious instances of non-compliance.

5 Discussion and Conclusion

Like all hierarchical organizations, the judicial hierarchy is replete with informational asymmetries. Our model puts these asymmetries front and center to understand how judges and external actors can simultaneously help higher courts with limited resources decide which cases to review *ex post*, and affect compliance by a lower court *ex ante*. Our main insights are twofold. First, we show that informativeness of a fire alarm to a higher court is

decreasing in its frequency; intuitively, the signal of a whistleblower who saves his warnings for more egregious instances of non-compliance will be more useful to a higher court, *ex post*, because it seeks information that helps it conserve scarce resources. Second, we show that the decreasing impact of whistleblowing, *ex post*, can eventually result in less compliance by a lower court *ex ante*. Our results thus illustrate the importance of identifying and connecting the *ex ante* and *ex post* effects of potential whistleblowers. This observation surely applies to organizations more generally, and further theoretical and empirical analysis would add to our understanding of how whistleblowers influence decision making in hierarchical institutions.

Once we consider the fact that higher courts oversee a number of lower courts and potential whistleblowers with varying preferences and incentives, our model helps us to understand a set of key empirical regularities about the federal judicial hierarchy. First, considering judges themselves as potential whistleblowers, our model illustrates how ideological heterogeneity and the institution of dissent interact to create panel effects on the Courts of Appeals. In particular, the model provides an internally consistent explanation for why potential dissenters can sometimes change the votes of their colleagues, but other times fail to do so and must act on their threat to dissent.

Second, when dissents or other whistles are blown, they do increase the likelihood of discretionary review, but review is far from guaranteed. For instance, the presence of a dissent or amicus brief increases the likelihood that the Supreme Court will review a case, but many such cases are nevertheless not reviewed. This, too, is consistent with our model—even when the higher court is sure that it disagrees with a decision, in some cases it will not find it worthwhile to review the decision.

Third, the effectiveness of the *threat* of dissent is a function of how often a potential whistleblower is actually willing to dissent. Thus, our model also illuminates why interest groups and the solicitor general are more effective at getting the Supreme Court to review lower court cases, relative to judges themselves. Because both interest groups' amici briefs at the certiorari stage and petitions by the solicitor general are relatively rare (compared to

the hundreds of dissents issued by Courts of Appeals judges in any given year), such signals are likely to be highly informative to the Supreme Court.¹³

Fourth, our model provides a rationale for the relative differences in the cost of sounding a fire alarm across legal actors. In the Appendix to the paper (section A-2) we present an institutional design analysis in which we characterize how the parameters in the whistleblower’s utility function—including the cost of dissent and its share of the reversal sanction—would be chosen to maximize the impact of whistleblowing on compliance. We show that preserving the informational quality of dissents from external whistleblowers who do not suffer any reversal sanction requires higher dissent costs than preserving the informational quality of internal whistleblowers who do suffer such costs. This result helps us to understand why sounding fire alarms is costlier for actors who are not judges, like interest groups filing amici briefs, than for judges themselves (i.e. to write dissents).

Finally, taking a step back, most decisions by lower courts are *not* accompanied by fire alarms. For example, the dissent rate on the Courts of Appeals is usually less than 10%. Our model suggests an explanation for this—fire alarms are only observed when the whistleblower’s excessive willingness to sound them diminishes their effectiveness so thoroughly that a lower court is willing to risk triggering them.

In addition to unifying existing empirical patterns, our model also generates non-obvious empirical implications about the relationship between judicial preferences, dissents, and higher court review. Existing work on strategic dissents on the Courts of Appeals theorizes that the likelihood of a dissent on a three-judge panel should be an increasing function of the distance between the potential whistleblower and the reviewing court; a perfect ally should be most likely to dissent (Hettinger, Lindquist and Martinek 2004, Blackstone and Collins 2011). However, as seen in Figure 3B, our model predicts that the likelihood of a dissent should continue to increase as the whistleblower becomes even farther away from the lower court majority than the higher court, *ceteris paribus*. These divergent predictions could be

¹³To be sure, the Solicitor General’s office is also so effective in triggering review because of its skilled lawyers and repeated interactions with the justices (Bailey, Kamoie and Maltzman 2005).

adjudicated with existing databases of judicial decision making (e.g. the Songer database) and current measures of judicial ideology (e.g. Giles-Hettinger-Peppers (2001) scores).

With respect to higher court review, the related empirical prediction is that the likelihood of higher court review following a dissent should be *decreasing* in the preference extremity of a whistleblower, *ceteris paribus* (see Figure 3C). In other words, a conservative higher court should be more likely to review a liberal decision accompanied by a dissent from a more liberal judge than from a more conservative judge. While existing studies of discretionary review have emphasized the importance of dissent (see e.g. Perry 1991, Caldeira, Wright and Zorn 1999), the importance of *who* is doing the dissenting has been under appreciated. Empirical tests along these lines would add more to our understanding of how the influence of whistleblowing extends throughout the judicial hierarchy.

Finally, our predictions about the relationship between the preferences of a whistleblower and compliance are quite subtle. As seen in Figure 4A, there is a non-monotonic relationship between the ideology of a potential whistleblower and lower court compliance. While measuring compliance is difficult, our model can be tested using the *votes* of a lower court (combined with measures of preferences of judges across the hierarchy). Our model suggests that as a whistleblower moves farther away from the lower court majority in the direction of a more conservative (liberal) higher court, the likelihood of a conservative (liberal) vote by the majority should first increase, but then eventually decrease. It is unlikely that this prediction would emerge outside a formal analysis of whistleblowing in the judicial hierarchy.

References

- Bailey, Michael A., Brian Kamoie and Forrest Maltzman. 2005. "Signals from the Tenth Justice: The Political Role of the Solicitor General in Supreme Court Decision Making." *American Journal of Political Science* 49(1):72–85.
- Baker, Scott and Claudio Mezzetti. 2012. "A Theory of Rational Jurisprudence." *Journal of Political Economy* 120(3):513–551.
- Bendor, Jonathan, Ami Glazer and Thomas Hammond. 2001. "Theories of Delegation." *Annual Review of Political Science* 2:235–69.
- Blackstone, Bethany and Paul M. Jr. Collins. 2011. "Strategy, Certainty, and the Decision

- to Dissent on the U.S. Courts of Appeals.” Paper presented at the 2011 annual meeting of the Midwest Political Science Association.
- Caldeira, Greg A., John R. Wright and Christopher J. W. Zorn. 1999. “Sophisticated Voting and Gate-Keeping in the Supreme Court.” *Journal of Law Economics & Organization* 15:549–72.
- Caldeira, Gregory A. and John R. Wright. 1988. “Organized Interests and Agenda Setting in the U.S. Supreme Court.” *The American Political Science Review* 82(4):1109–1127.
- Cameron, Charles M., Jeffrey A. Segal and Donald R. Songer. 2000. “Strategic Auditing in a Political Hierarchy: An Informational Model of the Supreme Court’s Certiorari Decisions.” *American Political Science Review* 94:101–16.
- Cameron, Charles M and Lewis A Kornhauser. 2010. “Adjudication Equilibria.” Princeton University working paper.
- Carrubba, Cliff, Barry Friedman, Andrew Martin and Georg Vanberg. 2012. “Who Controls the Content of Supreme Court Opinions?” *American Journal of Political Science* 56(2):400–12.
- Carrubba, Clifford J. and Tom S. Clark. 2012. “Rule Creation in a Political Hierarchy.” *American Political Science Review* 106(3):622–43.
- Cross, Frank B. and Emerson H. Tiller. 1998. “Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals.” *Yale Law Journal* 107:2155–76.
- Daughety, Andrew F. and Jennifer F. Reinganum. 2006. “Speaking Up: A Model of Judicial Dissent and Discretionary Review.” *Supreme Court Economic Review* 1:1–41.
- Epstein, David and Sharyn O’Halloran. 1995. “A Theory of Strategic Oversight: Congress, Lobbyists, and the Bureaucracy.” *Journal of Law, Economics, & Organization* 11:227–55.
- Fischman, Joshua B. 2011. “Estimating Preferences of Appellate Judges: A Model of Consensus Voting.” *Journal of Law and Economics* 54:781–809.
- Gailmard, Sean and John Patty. 2013. “Stovepiping.” *Journal of Theoretical Politics* 25:388–411.
- Gailmard, Sean and John W Patty. 2012. “Formal Models of Bureaucracy.” *Annual Review of Political Science* 15:353–377.
- George, Tracey E. 1999. “The Dynamics and Determinants of the Decision to Grant En Banc Review.” *Washington Law Review* 74:213–74.
- Giles, Micheal W., Thomas G. Walker and Christopher Zorn. 2006. “Setting a Judicial Agenda: The Decision to Grant En Banc Review in the U.S. Courts of Appeals.” *Journal of Politics* 68(4):852–66.

- Giles, Micheal W., Virginia A. Hettinger and Todd Peppers. 2001. "Picking Federal Judges: A Note on Policy and Partisan Selection Agendas." *Political Research Quarterly* 54:623–41.
- Ginsburg, Ruth B. 1990. "Remarks on Writing Separately." *Washington Law Review* 65:133–150.
- Hettinger, Virginia A., Stefanie A. Lindquist and Wendy L. Martinek. 2004. "Comparing Attitudinal and Strategic Accounts of Dissenting Behavior on the U.S. Courts of Appeals." *American Journal of Political Science* 48:123–37.
- Hopenhayn, Hugo and Susanne Lohmann. 1996. "Fire-Alarm Signals and the Political Oversight of Regulatory Agencies." *Journal of Law, Economics & Organization* 12(1):196–213.
- Kastellec, Jonathan P. 2007. "Panel Composition and Judicial Compliance on the U.S. Courts of Appeals." *Journal of Law, Economics & Organization* 23(2):421–441.
- Kornhauser, Lewis A. 1995. "Adjudication by a Resource-Constrained Team: Hierarchy and Precedent in a Judicial System." *Southern California Law Review* 68:1605–29.
- Lax, Jeffrey R. 2003. "Certiorari and Compliance in the Judicial Hierarchy: Discretion, Reputation and the Rule of Four." *Journal of Theoretical Politics* 15:61–86.
- Llewellyn, Karl N, Paul Gewirtz and Michael Ansaldi. 1988. "The Case Law System in America." *Columbia Law Review* 88(5):989–1020.
- Lupia, Arthur and Mathew D. McCubbins. 1994. "Learning from oversight: Fire alarms and police patrols reconstructed." *Journal of Law, Economics, & Organization* 10(1):96–125.
- McCubbins, Mathew D. and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28(1):165–79.
- McNollgast. 1995. "Politics and the Courts: A Positive Theory of Judicial Doctrine and the Rule of Law." *Southern California Law Review* 68:1631–89.
- Miller, Nathan H. 2009. "Strategic Leniency and Cartel Enforcement." *The American Economic Review* 99:750–768.
- Murphy, Walter. 1964. *Elements of Judicial Strategy*. Chicago: University of Chicago Press.
- Perry, Jr., H.W. 1991. *Deciding to Decide: Agenda Setting in the United States Supreme Court*. Cambridge: Harvard University Press.
- Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy* 111(5):929–58.
- Revesz, Richard L. 1997. "Environmental Regulation, Ideology and the D.C. Circuit." *Virginia Law Review* 83:1717–72.
- Spagnolo, Giancarlo. 2008. "Leniency and Whistleblowers in Antitrust." *Handbook of Antitrust Economics* pp. 259–304.

- Spitzer, Matthew and Eric Talley. 2000. "Judicial Auditing." *Journal of Legal Studies* 29:649–83.
- Spitzer, Matthew and Eric Talley. 2013. "Left, Right, and Center: Strategic Information Acquisition and Diversity in Judicial Panels." *Journal of Law Economics & Organization* 29:638–80.
- Sunstein, Cass R., David Schkade, Lisa M. Ellman and Andres Sawicki. 2006. *Are Judges Political? An Empirical Analysis of the Federal Judiciary*. Washington, D.C.: Brookings Institution Press.
- Takáts, Előd. 2011. "A Theory of 'Crying Wolf': The economics of money laundering enforcement." *Journal of Law, Economics, & Organization* 27(1):32–78.
- Tanenhaus, Joseph, Marvin Schick, Matthew Muraskin and Daniel Rosen. 1963. The Supreme Court's Certiorari Decisions: Cue Theory. In *Judicial Decision Making*, ed. Glendon Schubert. Glencoe, IL: Free Press pp. 111–32.
- Ting, Michael M. 2008. "Whistleblowing." *American Political Science Review* 102(2):249–67.

Appendix

A-1 Proofs

We first state and prove the following Lemma employed in many of the subsequent proofs.

Lemma 9 *The function $G(\phi_d(c_L, c_W)) - G(\phi_{nd}(c_L, c_W))$ is decreasing in c_W when $F(\cdot)$ and $G(\cdot)$ are uniform.*

Proof: With $G(\cdot)$ uniform it suffices to show $\phi_d(c_L, c_W) - \phi_{nd}(c_L, c_W)$ is decreasing in c_W . The function $\phi_d(c_L, c_W)$ may be rewritten as

$$\frac{\phi_{nd}(c_L, c_W)}{P(x \in [c_W, H] | x > c_W)} + (E[x | x \in [c_W, H]] - E[x | x \in [c_L, c_W]]),$$

and with substitution and algebra $\phi_d(c_L, c_W) - \phi_{nd}(c_L, c_W)$ may be rewritten as

$$P(x > H | x > c_W) \cdot (H - E[x | x \in [c_W, H]]) + (E[x | x \in [c_W, H]] - E[x | x \in [c_L, c_W]]).$$

Now for $F(\cdot)$ uniform $H - E[x | x \in [c_W, H]] = \left(\frac{\bar{x}}{2}\right) \cdot P(x \in [c_W, H])$. Substituting back in and simplifying we then have that $\phi_d(c_L, c_W) - \phi_{nd}(c_L, c_W) =$

$$\begin{aligned} \left(\frac{\bar{x}}{2}\right) \cdot P(x > H) (1 - P(x > H | x > c_W)) + (E[x | x \in [c_W, H]] - E[x | x \in [c_L, c_W]]) \text{ which } \rightarrow \\ \frac{\partial}{\partial c_W} (\phi_d(c_L, c_W) - \phi_{nd}(c_L, c_W)) &= -\left(\frac{\bar{x}}{2}\right) \cdot P(x > H) \frac{f(c_W)}{P(x > c_W)} P(x > H | x > c_W) \\ &\quad + \frac{\partial}{\partial c_W} (E[x | x \in [c_W, H]] - E[x | x \in [c_L, c_W]]) \\ &= -\left(\frac{\bar{x}}{2}\right) \cdot f(c_W) \cdot [P(x > H | x > c_W)]^2 < 0 \end{aligned}$$

since $\frac{\partial}{\partial c_W} (E[x | x \in [c_W, H]]) = \frac{\partial}{\partial c_W} (E[x | x \in [c_L, c_W]])$ for the uniform. ■

Proof of Lemma 1 As described in the beginning of Section 3 and footnote 4, in the main text we restrict attention to equilibria where dissent increases the probability of review – in other words, where $\phi_d > \phi_{nd}$. However, there sometimes exists a second more fragile class of equilibria with a different structure, in which “dissent” decreases the probability of review (i.e. $\phi_{nd} > \phi_d$) because it signals that noncompliance occurred but that it was relatively minor. Such “dissents” are more easily interpreted as “concurrences.” Below we prove a more

general statement about the form of all equilibria; Lemma 1 is a straightforward corollary of this more general statement.

Lemma 10 *All equilibria are in **cutpoint strategies** $(c_L^*, c_W^*, \phi_s^*, \phi_{ns}^*)$ with $c_L^* \in [L, H]$ and $c_W^* < H$. There are two types of equilibria.*

- In a **dissent** equilibrium,
 - The lower court L rules liberally for $x \geq c_L^*$ and conservatively otherwise.
 - The potential whistleblower W never dissents following a conservative ruling, and issues a costly dissent following a liberal ruling whenever the facts are sufficiently conservative ($x \leq c_W^*$).
 - The higher court H never reviews conservative rulings, and sometimes reviews liberal rulings. Specifically, he reviews a liberal ruling i.f.f. $k \leq \phi_s^*$ when W dissents, and i.f.f. $k \leq \phi_{ns}^* < \phi_s^*$ when W does not dissent.

- In a **concurrence** equilibrium,
 - The lower court L rules liberally for $x \geq c_L^*$ and conservatively otherwise.
 - The potential whistleblower W never concurs with a conservative ruling, and issues a costly concurrence following a liberal ruling when the facts are in $x \in [c_W^*, H]$.
 - The higher court H never reviews conservative rulings, and sometimes reviews liberal rulings. Specifically, he reviews a liberal ruling i.f.f. $k \leq \phi_{ns}^*$ when W fails to concur, and i.f.f. $k \leq \phi_s^* < \phi_{ns}^*$ when W concurs. ■

Proof: The interpretation of the costly signal is dependent on equilibrium. Specifically, it can either signal that the case facts are more conservative – in which case it is interpreted as a *dissent* – or it can signal that the case facts are more liberal – in which case it is interpreted as a *concurrence*. In the former instance it raises the probability of review, while

in the latter instance it lowers it. Thus, we denote the whistleblower's actions using the agnostic label $j \in \{s, ns\}$ – that is, she either issued the costly signal or she did not.

When H is called to play he is the final mover, and the history $h \in \{lib, con\} \times \{s, ns\}$ can take four possible values. For each history he calculates a *net gain* from review that is derived using Bayes' rule and the other players' strategies – denote this net gain $\phi_{i,j}$ where i denotes the ruling and j denotes the signal value. Because H is the last mover, his best-response takes the form of a cutpoint for each history – he reviews i.f.f $k < \phi_{i,j}$, where k is the cost of review.

Now consider the whistleblower W . If faced with a compliant ruling of either *lib* or *con* she will never send the costly signal, since the ruling will stand whether or not H reviews. Now suppose she is faced with a noncompliant ruling of *lib*. If she issues the costly signal then she will pay an up front cost of d . If the signal raises the chance of review ($\phi_{l,s} > \phi_{l,ns}$) then her net gain will be

$$(G(\phi_{l,s}) - G(\phi_{l,ns})) \cdot ((W - x) - \alpha\varepsilon),$$

since whenever $k \in (\phi_{l,s}, \phi_{l,ns})$ the signal results in a review and a reversal that otherwise would not have occurred. If the signal lowers the chance of review ($\phi_{l,s} < \phi_{l,ns}$) then her net gain will be

$$(G(\phi_{l,ns}) - G(\phi_{l,s})) \cdot ((x - W) + \alpha\varepsilon),$$

since whenever $k \in (\phi_{l,ns}, \phi_{l,s})$ the signal *prevents* a review and reversal that would otherwise have occurred. Recalling that she will never dissent on a compliant liberal ruling, her best response is either to signal when

$$x < \min \left\{ (W - \alpha\varepsilon) - \frac{d}{G(\phi_{l,s}) - G(\phi_{l,ns})}, H \right\}$$

if the signal increases review, or to signal when

$$x \in \left[(W - \alpha\varepsilon) + \frac{d}{G(\phi_{l,s}) - G(\phi_{l,ns})}, H \right]$$

if the signal decreases review. Thus, her strategy takes the desired forms.

Now consider the lower court L . It is strictly dominant to rule compliantly when it and the higher court agree ($x < L$ or $x > H$) since this ensures its desired outcome and there is no chance of being reversed (even upon review). For cases within $x \in [L, H]$ it is always the case that ruling liberally elicits a higher probability of review when $x < c_W^*$ than when $x > c_W^*$. If $\phi_s > \phi_{ns}$ then W signals when $x < c_W^*$, (thereby raising the probability of review) and if $\phi_{ns} > \phi_s$ then W fails to signal when $x < c_W^*$ (also raising the probability of review). Consequently, L must also play a cutpoint strategy c_L^* ; if it is unwilling to comply on some x then it is also unwilling to comply on some $x' > x$ where the benefits of the liberal outcome are greater and the probability of review is (weakly) lower.

Finally, because L always rules *lib* when $x > H$, any conservative ruling must be compliant. Since W never signals on a compliant ruling, PBE requires that $\phi_{c,ns} = 0$; that is, H evaluates the net gain of reviewing a conservative ruling without a signal to be 0 and never reviews it. A conservative ruling accompanied by a costly signal is off-path – in PBE $\phi_{c,s}$ is unrestricted, and the value will generate some off-path best response behavior for the whistleblower when she observes a noncompliant conservative ruling ($x > H$). However, choosing these pairs arbitrarily does not perturb equilibrium because ruling *lib* is strictly dominant for L whenever $x > H$ and a conservative ruling would be noncompliant; thus we leave these values unspecified. This completes the proof. ■

Proof of Lemmas 2 – 5 Lemmas 2 – 4 follow immediately from the in-text analysis. The necessary and sufficient condition for cutpoints $(c_L^*, c_W^*, \phi_{nd}^*, \phi_d^*)$ to be an equilibrium in Lemma 5 is a straightforward assembly of the best-response characterizations in the preceding Lemmas; that is, strategies are an equilibrium if and only if every player is best-responding down every path of play given the strategies of the other players. A more explicit statement of the assembled necessary and sufficient conditions is included below for clarity.

1. $\phi_d^* = \phi_d(c_L, \max\{c_W, c_L\})$ and $\phi_{nd}^* = \phi_{nd}(c_L, \max\{c_W, c_L\})$

(higher court best-response)

$$2. c_W^* = \min \{c_W(\phi_d^*, \phi_{nd}^*), H\}$$

(whistleblower best-response)

$$3. c_L^* = c_L(c_W^*, \phi_d^*, \phi_{nd}^*)$$

(lower court best-response). ■

Proof of Proposition 1 The proof proceeds in two parts. First, we show that the benchmarks \underline{c}_L and \bar{c}_L exist, are unique, and satisfy $L < \underline{c}_L < \bar{c}_L < H$. Second, we prove the main body of the statement.

Part 1

The cutpoint \underline{c}_L solves $c_L = x^*(G(\phi_{nd}(c_L, c_L)))$ and the cutpoint \bar{c}_L solves $c_L = x^*(G(\phi_d(c_L, c_L)))$. Using the definitions of $\phi_d(\cdot)$, $\phi_{nd}(\cdot)$, $x^*(\cdot)$, and that $G(0) = 0$, it is easily verified that the right hand sides of both equalities are (1) greater than L when $c_L = L$, (2) equal to L when $c_L = H$, and (3) strictly decreasing in c_L . Thus, both have a unique solution interior to (L, H) . Finally, to see that $\bar{c}_L > \underline{c}_L$ it suffices to observe that $\phi_d(c_L, c_L) > \phi_{nd}(c_L, c_L) \forall c_L$ and $x^*(G(\phi))$ is decreasing in ϕ .

Although the following is inessential to the proof, we now also briefly explain why \underline{c}_L and \bar{c}_L are the unique equilibrium levels of compliance in the no information and complete information 2 player games, respectively.

In the no information game absent the whistleblower, the higher court will use a single threshold ϕ for reviewing a liberal disposition, and this threshold must equal the expected benefit of review given his beliefs about the lower court's behavior. Applying the analysis in Section 3.3, the lower court must use a compliance cutpoint $c_L = x^*(G(\phi))$ in a best response. Given a cutpoint strategy by the lower court, the higher court must believe upon observing a liberal disposition that the case facts are $x > c_L$; it is easily verified that the net benefit of review under these circumstances is equal to $\phi_{nd}(c_L, c_L)$. Combining these two best response conditions yields the equilibrium condition $c_L = x^*(G(\phi_{nd}(c_L, c_L)))$.

In the complete information game the higher court observes the case facts to be x . Her expected benefit of reviewing and reversing the case, should the lower court rule noncompliantly, is equal to $H - x = \phi_d(x, x)$. Thus, she will review and reverse a noncompliant liberal ruling with probability $G(\phi_d(x, x))$. The lower court's net gain from noncompliance is $(1 - G(\phi_d(x, x)))(x - L)$ and the cost is $G(\phi_d(x, x)) \cdot \varepsilon$. The net gain is increasing and equal to 0 at $x = L$, and the cost is decreasing and equal to 0 at $x = H$; hence there is a unique interior cutpoint \hat{c}_L below which the lower court will comply and above which she will not, which satisfies

$$(1 - G(\phi_d(\hat{c}_L, \hat{c}_L)))(\hat{c}_L - L) = G(\phi_d(\hat{c}_L, \hat{c}_L)) \cdot \varepsilon \iff \hat{c}_L = x^*(G(\phi_d(\hat{c}_L, \hat{c}_L))),$$

which is the definition of \bar{c}_L

Part 2

We seek to characterize a *partial equilibrium* $(c_L^*, \phi_{nd}^*, \phi_d^*)$ where the lower and higher courts are best responding to each other and the whistleblower, and the whistleblower's strategy is to use a dissent cutpoint of c_W . In other words, we seek values of $(c_L^*, \phi_{nd}^*, \phi_d^*)$ that jointly satisfy Lemmas 2 and 4 given c_W .

By substituting the best-response conditions for the higher court into the best response condition for the lower court, we derive the following necessary and sufficient condition for existence of a partial equilibrium with compliance level c_L :

$$c_L = \min \{x^*(G(\phi_d(c_L, \max\{c_W, c_L\}))), \max\{x^*(G(\phi_{nd}(c_L, \max\{c_W, c_L\}))), c_W\}\} \quad (7)$$

The right hand side of the above is a function of c_L and c_W , and we henceforth denote it $\hat{c}_L(c_L; c_W)$. Intuitively, $\hat{c}_L(c_L; c_W)$ is the lower court's best response cutpoint when the higher court *believes it* to be using cutpoint c_L , and everybody believes the whistleblower to be using cutpoint c_W .¹⁴ A partial equilibrium level of compliance is a fixed point of this

¹⁴Note that c_W is a complete contingent description of how W would behave after a liberal ruling on any case $x \in X$. L cannot "change c_W " off-equilibrium path. Rather, L 's ruling, combined with the case facts, determine whether or not W dissents *based on* c_W . If, for example, W and L 's strategies are described by cutpoints $L < c_L < c_W < H$, then W 's strategy specifies precisely what would happen if L were to go "off-path" by ruling liberally on a case $x \in [L, c_L]$ – it would trigger a dissent.

function.

We now show that for every c_W there exists a *unique* value of c_L satisfying the equality in (7), and that value is equal to the cutpoint $c_L^*(c_L)$ described in the Proposition. First, using the definitions of $\phi_d(\cdot)$, $\phi_{nd}(\cdot)$, $x^*(\cdot)$, and that $G(0) = 0$, the following facts about the r.h.s. of the equality are easily verified: (1) it is weakly decreasing in c_L (since it is the middle value of three weakly decreasing functions), 2) $c_L^*(L, c_W) > L$, and 3) $c_L^*(H, c_W) = L$. This establishes that there is a unique solution interior to (L, H) .

For the next steps also recall that $\underline{c}_L = x^*(G(\phi_{nd}(\underline{c}_L, \underline{c}_L)))$ and $\bar{c}_L = x^*(G(\phi_d(\bar{c}_L, \bar{c}_L)))$.

Region 1: To see that $\underline{c}_L = \hat{c}_L(\underline{c}_L; c_W)$ when $c_W < \underline{c}_L$, note that the latter implies

$$\begin{aligned} \hat{c}_L(\underline{c}_L, c_W) &= \min \{x^*(G(\phi_d(\underline{c}_L, \underline{c}_L))), \max \{x^*(G(\phi_{nd}(\underline{c}_L, \underline{c}_L))), c_W\}\} \\ &= \min \{x^*(G(\phi_d(\underline{c}_L, \underline{c}_L))), \max \{\underline{c}_L, c_W\}\} = \underline{c}_L \end{aligned}$$

Intuitively, when c_W is less than the cutpoint \underline{c}_L that the lower court would use absent the whistleblower, then the lower court's partial equilibrium compliance cutpoint is the same as absent W – the whistleblower never dissents on path, absent dissents the higher court draws the same inference as he would absent the whistleblower, and so the lower court complies to the same degree. In this case, the degree of compliance is constant in c_W and equal to \underline{c}_L , the probability of review after dissent is $G(\phi_d(\underline{c}_L, \underline{c}_L))$ and also constant, and the probability of dissent is 0 (since $c_W < \underline{c}_L$).

Region 2: To see that $c_L^*(c_W) = c_W \iff c_W = \hat{c}_L(c_W, c_W)$ when $c_W \in [\underline{c}_L, \bar{c}_L]$, note that the latter implies (from the definitions of \underline{c}_L and \bar{c}_L) that $c_W > x^*(G(\phi_{nd}(c_W, c_W)))$ and that $c_W < x^*(G(\phi_d(c_W, c_W)))$. Hence,

$$\hat{c}_L(c_W; c_W) = \min \{x^*(G(\phi_d(c_W, c_W))), \max \{x^*(G(\phi_{nd}(c_W, c_W))), c_W\}\} = c_W.$$

Intuitively, suppose $c_W \in [\underline{c}_L, \bar{c}_L]$ and the lower court were to comply exactly up to c_W . A dissent would perfectly signal that the case facts were at c_W , and thus L would want to comply for all $x < c_W$ since c_W is less than the cutpoint \bar{c}_L it would use if H were perfectly informed about the case facts. Conversely, the absence of dissent would signal that $x > c_W$;

since L would not comply on such cases if H drew the inference that $x \in [\underline{c}_L, H]$, she also would not comply when H draws the weaker inference that $x \in [c_W, H]$. Consequently, L 's best response cutpoint is exactly at $\hat{c}_L(c_W) = c_W$ and we have a partial equilibrium. In this region, compliance is clearly increasing since it is equal to c_W , and the probability of review after dissent is $G(\phi_d(c_W, c_W)) = G(H - c_W)$ which is also decreasing in c_W .

Region 3: Suppose that $c_W > \bar{c}_L$ and denote $c_L^d(c_W)$ as the value of c_L that solves $c_L = x^*(G(\phi_d(c_L, c_W)))$. It is easily verified using steps identical to those for \underline{c}_L and \bar{c}_L that $c_L^d(c_W)$ is unique, well defined, and in (L, H) . We now wish to prove that $c_W > \bar{c}_L \rightarrow c_L^d(c_W) = \hat{c}_L(c_L^d(c_W), c_W)$, meaning that the partial equilibrium compliance cutpoint is exactly $c_L^d(c_W)$. First, note that $\bar{c}_L = x^*(G(\phi_d(\bar{c}_L, \bar{c}_L))) \rightarrow \bar{c}_L > x^*(G(\phi_d(\bar{c}_L, c_W)))$ for $c_W > \bar{c}_L$ (since $x^*(G(\phi_d(c_L, c_W)))$ is decreasing in c_W). This then implies that the solution to $c_L^d(c_W) = x^*(G(\phi_d(c_L^d(c_W), c_W)))$ is $< \bar{c}_L < c_W$. Now from the definition of $\hat{c}_L(c_L, c_W)$, the lower court's best response cutpoint for any (c_L, c_W) such that $c_W > x^*(G(\phi_d(c_L, \max\{c_W, c_L\})))$ is equal to $x^*(G(\phi_d(c_L, \max\{c_W, c_L\})))$; thus, the best response cutpoint to $(c_L^d(c_W), c_W)$ is $c_L^d(c_W)$ and we have a partial equilibrium.

To see the comparative statics, first note $c_L^d(c_W) = x^*(G(\phi_d(c_L^d(c_W), c_W)))$ is decreasing in c_W since $\phi_d(\cdot)$ is decreasing in c_W , $G(\phi)$ is increasing in ϕ , and $x^*(q)$ is increasing in q . The probability of dissent $F(c_W) - F(c_L^d(c_W))$ is then increasing in c_W since $c_L^d(c_W)$ is decreasing in c_W . Finally, to see that the probability of review given dissent $G(\phi_d(c_L^d(c_W), c_W))$ is decreasing in c_W , implicitly differentiate the definition to get,

$$\frac{\partial}{\partial c_W} (c_L^d(c_W)) = \frac{\partial x^*(G(\phi_d(c_L^d(c_W), c_W)))}{\partial q} \cdot \frac{\partial}{\partial c_W} (G(\phi_d(c_L^d(c_W), c_W)))$$

Since $\frac{\partial x^*(\cdot)}{\partial q} > 0$, $\frac{\partial}{\partial c_W} (G(\phi_d(c_L^d(c_W), c_W)))$ inherits the sign of $\frac{\partial}{\partial c_W} (c_L^d(c_W))$ which as previously shown is negative. ■

Proof of Proposition 2

Part 1 - Equilibrium Characterization and Existence

A necessary and sufficient condition for a profile of cutpoints $(c_L^*, c_W^*, \phi_{nd}^*, \phi_d^*)$ to be an

equilibrium of the complete model is that they jointly satisfy Lemmas 2 – 4. By Proposition 1, if the whistleblower uses cutpoint c_W then equilibrium requires that $c_L^* = c_L^*(c_W)$ (which is uniquely defined) and thus that $\phi_d^* = \phi_d(c_L^*(c_W), \max\{c_L^*(c_W), c_W\})$ and $\phi_{nd}^* = \phi_{nd}(c_L^*(c_W), \max\{c_L^*(c_W), c_W\})$, which are also uniquely defined. Because the necessary values of the other players strategies in an equilibrium are uniquely pinned down for every c_W , we can substitute these values into the whistleblower's best response characterization in Lemma 3 to yield a necessary and sufficient condition for existence of an equilibrium with whistleblowing cutpoint c_W :

$$c_W = \min \{c_W(\phi_d(c_L^*(c_W), \max\{c_L^*(c_W), c_W\}), \phi_{nd}(c_L^*(c_W), \max\{c_L^*(c_W), c_W\})), H\}.$$

Observe that the right hand side is a function of c_W alone, and we henceforth denote it $\hat{c}_W(c_W)$. Equilibrium values of c_W are fixed points of this function; the equilibrium condition in the main text is identical except with the definition of $c_W(\phi_d, \phi_{nd})$ substituted in. Intuitively, $\hat{c}_W(c_W)$ is the whistleblower's best response cutpoint when the lower and higher court *believe* her to be using cutpoint c_W , and play their corresponding partial equilibrium strategies.

Existence of an equilibrium whistleblowing cutpoint satisfying $c_W = \hat{c}_W(c_W)$ that is $\leq H$ (and hence an equilibrium of the complete model) then follows immediately from the fact that $\hat{c}_W(c_W) \leq H \forall c_W$.

Part 2 - Necessary and Sufficient Condition for Whistleblower Effects

It is helpful to more-explicitly write the definition of $\hat{c}_W(c_W)$ by substituting in the partial equilibrium values of the lower court's compliance $c_L^*(c_W)$. We have that

$$\hat{c}_W(c_W) = \begin{cases} \min \left\{ (W - \alpha\varepsilon) - \frac{d}{G(\phi_d(\underline{c}_L, \underline{c}_L)) - G(\phi_{nd}(\underline{c}_L, \underline{c}_L))}, H \right\} & \text{for } c_W \leq \underline{c}_L. \\ \min \left\{ (W - \alpha\varepsilon) - \frac{d}{G(\phi_d(c_W, c_W)) - G(\phi_{nd}(c_W, c_W))}, H \right\} & \text{for } c_W \in [\underline{c}_L, \bar{c}_L] \\ \min \left\{ (W - \alpha\varepsilon) - \frac{d}{G(\phi_d(c_L^*(c_W), c_W)) - G(\phi_{nd}(c_L^*(c_W), c_W))}, H \right\} & \text{for } c_W \geq \bar{c}_L \end{cases}$$

In the main proof we employ the following useful properties of $\hat{c}_W(c_W)$. First, it is constant for $c_W \leq \underline{c}_L$. Second, it is weakly decreasing for $c_W \in [\underline{c}_L, \bar{c}_L]$, and strictly decreasing if

$\hat{c}_W(c_W) < H$. Third, $\hat{c}_W(\underline{c}_L) \geq \hat{c}_W(c_W)$ for any $c_W > \underline{c}_L$.

To show these properties, first notice that c_W only affects $\hat{c}_W(c_W)$ through the denominator of the fraction in the first term – this is the probability that dissent is *pivotal* for review given cutpoint c_W . Now, the first property is immediate from the definition. To prove the second and third properties, we argue as an intermediate step that $G(\phi_d(c_L, c_W)) - G(\phi_{nd}(c_L, c_W)) > G(\phi_d(c'_L, c'_W)) - G(\phi_{nd}(c'_L, c'_W))$ for $c'_L > c_L$, $c'_W > c_W$, $c_L \leq c_W$ and $c'_L \leq c'_W$. This is because

$$\begin{aligned} G(\phi_d(c_L, c_W)) - G(\phi_{nd}(c_L, c_W)) &> G(\phi_d(c_L, c'_W)) - G(\phi_{nd}(c_L, c'_W)) \quad (\text{by Lemma 9}) \\ &> G(\phi_d(c'_L, c'_W)) - G(\phi_{nd}(c'_L, c'_W)) \quad (\text{by definitions}) \end{aligned}$$

This then implies both that $G(\phi_d(c_W, c_W)) - G(\phi_{nd}(c_W, c_W))$ is strictly decreasing for $c_W \in [\underline{c}_L, \bar{c}_L]$ (implying the second property) and that $G(\phi_d(\underline{c}_L, \underline{c}_L)) - G(\phi_{nd}(\underline{c}_L, \underline{c}_L)) > G(\phi_d(c_L^*(c_W), c_W)) - G(\phi_{nd}(c_L^*(c_W), c_W))$ for $c_W > \underline{c}_L$ (implying the third property).

We now proceed to the main proof. By the definition of $c_L^*(c_W)$, whistleblower effects – that is, greater compliance than \underline{c}_L – occur in an equilibrium if and only if the equilibrium whistleblowing cutpoint c_W^* is $> \underline{c}_L$. To show that the desired condition is necessary and sufficient, it then suffices to show that (1) if it fails all equilibrium whistleblowing cutpoints are $\leq \underline{c}_L$, and (2) if it holds there exists an equilibrium whistleblowing cutpoint $> \underline{c}_L$.

(Necessity). If the condition fails, then $H > \underline{c}_L > (W - \alpha\varepsilon) - \frac{d}{G(\phi_d(\underline{c}_L, \underline{c}_L)) - G(\phi_{nd}(\underline{c}_L, \underline{c}_L))}$ which implies that $\underline{c}_L > \hat{c}_W(\underline{c}_L)$. Since $\hat{c}_W(\underline{c}_L) \geq \hat{c}_W(c_W) \forall c_W > \hat{c}_W(\underline{c}_L)$, there are no equilibrium whistleblowing cutpoints greater than \underline{c}_L and hence no equilibria with whistleblower effects.

(Sufficiency) If the condition holds then $(W - \alpha\varepsilon) - \frac{d}{G(\phi_d(\underline{c}_L, \underline{c}_L)) - G(\phi_{nd}(\underline{c}_L, \underline{c}_L))} < \underline{c}_L < H$ which implies that $\underline{c}_L < \hat{c}_W(\underline{c}_L)$. Since $\hat{c}_W(\underline{c}_L)$ is constant for $c_W < \underline{c}_L$, this implies that all equilibrium whistleblowing cutpoints c_W^* are $> \underline{c}_L$. Consequently, all equilibria exhibit whistleblower effects (which is in fact stronger than the desired property). ■

Proof of Lemma 6 For the purposes of this proof it is helpful to explicitly express the dependence of the best response mapping on the whistleblower's parameters, i.e. $\hat{c}_W(c_W; W, d, \alpha)$.

Several substantively unimportant subtleties are now worth noting. First, the mapping from the parameter space to maximum equilibrium compliance $\tilde{c}_L(W, d, \alpha)$ is not necessarily continuous. Second, in the comparative statics for each parameter (W, d, α) there need not always be a region with partial and *strictly diminishing* whistleblower effects. Instead, the region with full whistleblower effects may jump to one with partial whistleblower effects where compliance is constant. Third, the regions may be truncated for the cost of dissent d because it cannot fall below 0;¹⁵ for example, if $W = L$ and $\alpha = 0$, then there would be no whistleblower effects in a dissent equilibrium for any feasible value of $d \geq 0$.

In this proof we formally describe the steps for the parameter W ; steps for d, α are the same except the order of the regions is reversed (and thus identical for $-d$ and $-\alpha$) with the understanding that the parameter space for d is truncated. The proof proceeds in three parts. First, we show that when there are multiple equilibria the compliance maximizing equilibrium is the one with the lowest c_W ; we denote this whistleblowing cutpoint $\tilde{c}_W(W, d, \alpha) = \min \{c_W : c_W = \hat{c}_W(c_W; W, d, \alpha)\}$. Hence, maximum equilibrium compliance $\tilde{c}_L(W, d, \alpha)$ is equal to the composite mapping $c_L^*(\tilde{c}_W(W, d, \alpha))$. Second, we prove several properties of $\tilde{c}_W(W, d, \alpha)$. Third, we apply parts 1 and 2 to show the desired result.

Part 1

We argue that when there are multiple equilibria, the compliance maximizing equilibrium is the one with the lowest c_W^* . Intuitively, this holds because to sustain a higher equilibrium whistleblowing cutpoint, the whistleblower’s probability of being pivotal must be higher, and since more whistleblowing reduces the probability of being pivotal *ceteris paribus*, this necessarily requires less compliance.

Recall from the proof of Proposition 2 that $\hat{c}_W(c_W; W, d, \alpha)$ is constant over $c_W \leq \underline{c}_L$ and larger at \underline{c}_L than at any $c_W > \underline{c}_L$. As a result, there cannot be multiple equilibria where one exhibits no whistleblower effects and others do—either there is a unique equilibrium with

¹⁵This is not an assumption but an observation; if dissent in the literal real-world sense were beneficial rather than costly, then choosing not to dissent would be the costly signal of “dissent” in the model.

no whistleblower effects ($c_W^* \leq \underline{c}_L$) or there are one or more equilibria all of which exhibit whistleblower effects ($c_W^* > \underline{c}_L$).

Now, if there are two equilibria with whistleblower effects $\hat{c}_W^* > c_W^* > \underline{c}_L$, then by definition,

$$\begin{aligned} c_W^* &= \min \left\{ (W - \alpha\varepsilon) - \frac{d}{G(\phi_d(c_L^*(c_W^*), c_W^*)) - G(\phi_{nd}(c_L^*(c_W^*), c_W^*))}, H \right\} \text{ and} \\ \hat{c}_W^* &= \min \left\{ (W - \alpha\varepsilon) - \frac{d}{G(\phi_d(c_L^*(\hat{c}_W^*), \hat{c}_W^*)) - G(\phi_{nd}(c_L^*(\hat{c}_W^*), \hat{c}_W^*))}, H \right\} \end{aligned}$$

This implies that,

$$G(\phi_d(c_L^*(\hat{c}_W^*), \hat{c}_W^*)) - G(\phi_{nd}(c_L^*(\hat{c}_W^*), \hat{c}_W^*)) > G(\phi_d(c_L^*(c_W^*), c_W^*)) - G(\phi_{nd}(c_L^*(c_W^*), c_W^*)),$$

which in turn could only be true if $c_L^*(\hat{c}_W^*) < c_L^*(c_W^*)$, since by Lemma 9 the difference $G(\phi_d(c_L, c_W)) - G(\phi_{nd}(c_L, c_W))$ is decreasing in c_W . This shows the desired property.

Part 2

We now show that $\tilde{c}_W(W, d, \alpha)$ satisfies the following three properties:

1. it is weakly increasing in W
2. for any value of $c_W^* \in [-\infty, \bar{c}_L]$, there \exists a unique W^* s.t. $\tilde{c}_W(W^*, d, \alpha) = c_W^*$
3. there $\exists W$ s.t. $\tilde{c}_W(W, d, \alpha) = H$

To see (1) consider two values of the whistleblower $W' > W$. By definition of $\tilde{c}_W(W, d, \alpha)$, any value of $c_W < \tilde{c}_W(W, d, \alpha)$ is also less than $\hat{c}_W(c_W; W, d, \alpha)$ (because it is less than the lowest fixed point). Since $\hat{c}_W(c_W; W, d, \alpha)$ is increasing in W , this furthermore implies that any value of $c_W < \tilde{c}_W(W, d, \alpha)$ is also less than $\hat{c}_W(c_W; W', d, \alpha)$. Thus, the lowest fixed point $\tilde{c}_W(W', d, \alpha)$ for W' must be \geq the lowest fixed point $\tilde{c}_W(W, d, \alpha)$ for W .

To see (2), it is easy to verify from the equilibrium definition in Proposition 2 that for any value of $c_W^* < H$ there is a *unique* W^* s.t. c_W^* is an equilibrium whistleblowing cutpoint. However, this is not enough to show that $\tilde{c}_W(W^*, d, \alpha) = c_W^*$, i.e., that c_W^* is the *lowest* equilibrium whistleblowing cutpoint for W^* . We now show this property must also hold for

any $c_W^* \leq \bar{c}_L$. To do so, it suffices to recall from the proof of Proposition 2 that the best-response mapping $\hat{c}_W(c_W; W, d, \alpha)$ is weakly decreasing for $c_W \leq \bar{c}_L$. Consequently, there is at most one fixed point $\leq \bar{c}_L$, so if such an equilibrium exists it must be the lowest one.

To see (3), observe that the probability dissent is pivotal $G(\phi_d(c_L^*(c_W), c_W)) - G(\phi_{nd}(c_L^*(c_W), c_W))$ is $> G(\phi_d(\bar{c}_L, H)) - G(\phi_{nd}(\bar{c}_L, H)) > 0$ by $c_L^*(c_W) \leq \bar{c}_L$ and Lemma 9. Thus, for any whistleblower W such that,

$$(W - \alpha\varepsilon) - \frac{d}{G(\phi_d(\bar{c}_L, H)) - G(\phi_{nd}(\bar{c}_L, H))} > H$$

the best response mapping $\hat{c}_W(c_W; W, d, \alpha)$ is also $> H$ for any c_W , and consequently the unique equilibrium involves whistleblowing cutpoint $c_W^* = H$.

Part 3

The properties proved in Part 2 jointly imply that (1) $\tilde{c}_W(W, d, \alpha)$ is first strictly increasing, continuous in W , and onto $[-\infty, \bar{c}_L]$ (2) beyond \bar{c}_L the function continues to increase (but potentially discontinuously) until it reaches H , and (3) it is constant thereafter. Consequently, maximum equilibrium compliance $c_L^*(\tilde{c}_W(W, d, \alpha))$ exhibits the regions as described mirroring the regions of c_W – first with no whistleblower effects, followed by continuously increasing whistleblower effects up to \bar{c}_L , followed by (potentially discontinuously decreasing) partial whistleblower effects, and finally constant and partial whistleblower effects when W is fully reporting all instances of noncompliance. Properties and analysis are identical for $(-\alpha, -d)$ except that for d the regions may be truncated from the top.

Proof of Lemma 7 The proof proceeds in two parts. First, we characterize H 's expected utility as a function of the whistleblower's cutpoint c_W in a partial equilibrium, as well as the derivative of that utility. Second, we use this analysis to prove the main results.

Part 1

In Region I of Proposition 1, H 's expected utility as a function of c_W is constant since $c_L^*(c_W) = \underline{c}_L$ and dissent is off path. In Regions II and III H 's complete expected utility,

taking into account his review costs, is the expression:

$$\begin{aligned} & \int_{-\infty}^{c_L^*(c_W)} \left(\frac{H-x}{2} \right) f(x) dx + \int_H^\infty \int_0^{\phi_{nd}(c_L^*(c_W), c_W)} -kg(k) dk f(x) dx \\ & + \int_{c_L^*(c_W)}^{c_W} \left(\int_0^{\phi_d(c_L^*(c_W), c_W)} \left(\left(\frac{H-x}{2} \right) - k \right) g(k) dk + \int_{\phi_d(c_L^*(c_W), c_W)}^\infty \left(\frac{x-H}{2} \right) g(k) dk \right) f(x) dx \\ & + \int_{c_W}^H \left(\int_0^{\phi_{nd}(c_L^*(c_W), c_W)} \left(\left(\frac{H-x}{2} \right) - k \right) g(k) dk + \int_{\phi_{nd}(c_L^*(c_W), c_W)}^\infty \left(\frac{x-H}{2} \right) g(k) dk \right) f(x) dx \end{aligned}$$

This expression has a simple and easily interpretable derivative in the whistleblower's cutpoint c_W which is derived using Leibniz rule and canceling:

$$\begin{aligned} & \left(\frac{\partial c_L^*(c_W)}{c_W} \right) f(c_L^*(c_W)) \left(\int_0^{\phi_d(c_L^*(c_W), c_W)} kg(k) + (1 - G(\phi_d(\cdot))) (H - c_L^*(c_W)) \right) \\ & + f(c_W) \left(\int_{\phi_{nd}(c_L^*(c_W), c_W)}^{\phi_d(c_L^*(c_W), c_W)} ((H - c_W) - k) g(k) dk \right). \end{aligned} \quad (8)$$

The first line is the net gain resulting from the change in L 's compliance behavior $c_L^*(c_W)$. It is the product of three subterms: 1) the density $f(c_L^*(c_W))$ of cases at the compliance cutpoint, 2) the marginal change $\frac{\partial c_L^*(c_W)}{c_W}$ in the compliance cutpoint, and 3) the marginal benefit $\int_0^{\phi_d(c_L^*(c_W), c_W)} kg(k) + (1 - G(\phi_d(\cdot))) (H - c_L^*(c_W))$ of switching from noncompliance to compliance at case $x = c_L^*(c_W)$. (This is because when $k < \phi_d(c_L^*(c_W), c_W)$ the outcome doesn't change but H saves the review cost, while when $k > \phi_d(c_L^*(c_W), c_W)$ H would not have reviewed either way but now gets a compliant outcome for free.)

The second line is the net gain or loss from the whistleblower sending the costly rather than free signal at case $x = c_W$, which results in H inferring that x is in $[c_L, c_W]$ rather than $[c_W, \bar{x}]$. This net gain is comprised of the density of cases $f(c_W)$ at the whistleblowing cutpoint, times the net benefit of obtaining the conservative outcome through a review when $k \in [\phi_{nd}(\cdot), \phi_d(\cdot)]$.

Part 2

We now show that the utility-maximizing whistleblowing cutpoint for H is strictly less than H and weakly greater than L ; we do so by showing that the derivative is < 0 at H and > 0 at $c_W \in (c_L, \bar{c}_L)$.

At $c_W = H$ we have $\left. \frac{\partial c_L^*(c_W)}{c_W} \right|_{c_W=H} < 0$ (so the first term is negative) and the second term reduces to $-f(c_W) \int_{\phi_{nd}(c_L^*(H), H)}^{\phi_d(c_L^*(H), H)} kg(k) dk < 0$. Intuitively, complete reporting of noncompliance is both costly in terms of compliance, and more reporting than H wants even absent the compliance effect.

For $c_W \in (\underline{c}_L, \bar{c}_L)$ we have $c_L^*(c_W) = c_W$ and the derivative simplifies to,

$$f(c_W) \left(\int_0^{\phi_{nd}(c_W, c_W)} kg(k) dk + \int_{\phi_{nd}(c_W, c_W)}^{\infty} (H - c_W) \right) > 0.$$

Intuitively, more whistleblowing is all gain since it converts the marginal case from one where the lower court is compliant only when reviewed, to one on which the lower court complies for sure. Thus, the utility maximizing cutpoint is $\geq \bar{c}_L$. ■

Proof of Lemma 8 Recall from the analysis in the proof of Lemma 7 that H 's preferences for changes in the whistleblower's cutpoint c_W involves a trade off between the *equilibrium compliance cost* of more whistleblowing against the *marginal informational benefit*.

From equation (8), this marginal informational benefit (henceforth "MIB") is equal to,

$$f(c_W) \left(\int_{\phi_{nd}(c_L^*(c_W), c_W)}^{\phi_d(c_L^*(c_W), c_W)} ((H - c_W) - k) g(k) dk \right)$$

The proof requires two substeps. First, we show that the MIB satisfies a single crossing property and is equal to 0 at some unique $c_W^{**} \in (\bar{c}_L, H)$. Second, we show that the equilibrium of the game where the whistleblower is a perfect agent who internalizes H 's review costs involves that whistleblower using cutpoint c_W^{**} , and the higher and lower courts jointly best responding exactly as in the baseline model. These two properties then jointly imply that at the unique equilibrium with a perfect agent, the whistleblower's cutpoint is strictly greater than the cutpoint maximizing H 's utility. The reason is that a necessary condition for some \hat{c}_W to maximize H 's utility (from eqn. 8) is for the MIB be equal to the marginal compliance cost. Since the marginal compliance cost is always strictly positive in Region III, at the utility maximizing \hat{c}_W the MIB must also be strictly positive, so \hat{c}_W must be strictly less than c_W^{**} .

Intuitively, a "clone" of H as whistleblower dissents too much because – lacking commit-

ment power and responding to her interim incentives – she only takes into account the MIB of more whistleblowing and not the marginal compliance cost. (A clone who could commit ex-ante to her whistleblowing behavior would indeed induce the optimum for H).

Part 1

From the proof in Lemma 7, recall that the MIB is positive in Region II and negative at $c_W = H$. In Region III it can be rewritten as,

$$\left(\frac{\phi_d(\cdot) - \phi_{nd}(\cdot)}{\bar{x}k} \right) \cdot \left((H - c_W) - \frac{\phi_d(c_L^*(c_W), c_W) + \phi_{nd}(c_L^*(c_W), c_W)}{2} \right)$$

To show this satisfies a single crossing property it suffices to show that the second term is decreasing in c_W , which can be written as,

$$\frac{1}{2} \left(((H - c_W) - \phi_d(c_L^*(c_W), c_W)) + ((H - c_W) - \phi_{nd}(c_L^*(c_W), c_W)) \right)$$

It is simple to verify that $(H - c_W) - \phi_d(c_L^*(c_W), c_W)$ is decreasing in c_W . Hence for the desired property it suffices to show that $(H - c_W) - \phi_{nd}(c_L^*(c_W), c_W)$ is also decreasing in c_W , which in turn holds if $\frac{\partial(\phi_{nd}(\cdot))}{\partial c_W} > -1$. Taking this derivative we have

$$\frac{\partial}{\partial c_W} \left(\frac{(H - c_W)^2}{2(\bar{x} - c_W)} \right) = -\frac{H - c_W}{\bar{x} - c_W} + \frac{(H - c_W)^2}{2(\bar{x} - c_W)^2},$$

which immediately shows the desired property because $\frac{H - c_W}{\bar{x} - c_W} < 1$.

Part 2

In the slightly modified game where W internalizes H 's review costs, H still uses cutpoint strategies as in the baseline model. When $\phi_d > \phi_{nd}$, W 's net benefit of dissenting on a liberal ruling on x – conditional on that dissent being pivotal for review (i.e. $k \in [\phi_{nd}, \phi_d]$) – is now modified to be equal to $((W - x) - \alpha\varepsilon) - E[k | k \in [\phi_{nd}, \phi_d]]$, because she internalizes k . The net cost of dissent is again d . Thus as in the baseline model W uses a cutpoint strategy of dissenting whenever x is less than the minimum of H and

$$(W - \alpha\varepsilon - E[k | k \in [\phi_{nd}, \phi_d]]) - \frac{d}{G(\phi_d) - G(\phi_{nd})}$$

This in turn implies that L uses a cutpoint strategy, and that the form of the equilibrium and the partial equilibrium conditions from Proposition 1 are unchanged.

Now when W is a perfect agent in the sense of preferences, her ideal cutpoint is $= H$, $\alpha = 0$, and $d = 0$. Thus, her best response cutpoint is $H - E[k | k \in [\phi_{nd}, \phi_d]]$, which must be equal to the other players' beliefs about it in equilibrium. Substituting in the partial equilibrium conditions implies that a necessary and sufficient condition for an equilibrium with whistleblower cutpoint c_W is then

$$\begin{aligned}
H - \frac{\phi_d(c_L^*(c_W), c_W) + \phi_{nd}(c_L^*(c_W), c_W)}{2} &= c_W \\
\iff (H - c_W) - \frac{\phi_d(c_L^*(c_W), c_W) + \phi_{nd}(c_L^*(c_W), c_W)}{2} &= 0
\end{aligned}$$

which is equivalent to the condition for the MIB to be $= 0$. ■

A-2 Institutional Design Analysis

In this supplemental analysis to the main text, we consider how an institutional designer would select the parameters in the whistleblower’s utility function—her conservatism W , the cost of dissent d , and her share of the sanction α —to maximize compliance. Compliance is maximized when the whistleblower’s payoffs are calibrated so that dissent is attractive, but not too attractive. Specifically, she must be willing to dissent *exactly up* to the intermediate “limit to compliance” (\bar{c}_L) derived in Proposition 1 and no further. Any less dissenting and compliance gains are foregone because the threat of dissent can induce more compliance. Any more dissenting is counterproductive due to the negative equilibrium effect of reducing the impact of dissent. The condition for dissenting precisely up to the limit \bar{c}_L to constitute an equilibrium is stated in the following proposition.

Proposition 3 *Holding other parameters of the model fixed, compliance by the lower court is maximized when W , d , and α are jointly chosen so that the following equality holds:*

$$W - \alpha\varepsilon = \bar{c}_L + \frac{d}{G(\phi_d(\bar{c}_L, \bar{c}_L)) - G(\phi_{nd}(\bar{c}_L, \bar{c}_L))}. \quad (9)$$

Using Proposition 3, we can extract a number of substantively interesting results about compliance-maximizing institutional design. We first consider how an institutional designer with power to choose only one of the cost of dissent d , the whistleblower’s share of the sanction α , or the whistleblower’s indifference point W , would change the parameter of interest in response to changes in one of the other two.

Corollary 2 *A compliance-maximizing institutional designer choosing d , α , or W would:*

- *lower the cost of dissent d if the whistleblower became more liberal or if her share of the sanction α increased;*¹⁶
- *choose a more conservative whistleblower if the cost of dissent d increased or the whistleblower’s share of the sanction α increased;*

¹⁶Unless d were already 0, in which case she would leave it there.

- *decrease the whistleblower’s share of the sanction α if the cost of dissent d increased or the whistleblower became more liberal.*

Intuitively, these comparative statics arise from the fact that decreasing the cost of dissent d , decreasing the sanction share α , and increasing whistleblower’s conservatism W , are *substitutable* ways of increasing the whistleblower’s willingness to dissent. Since compliance maximization requires intermediate whistleblowing, a compliance-maximizing institutional designer should respond to an increase in the whistleblower’s intrinsic willingness to dissent by tamping down on the incentive to dissent—either by increasing the whistleblower’s sanction share, increasing the cost of dissent, or by choosing a more liberal whistleblower.

Next, we consider how an institutional designer choosing one of d , α , or W would change that parameter in response to changes in the lower court’s willingness to comply, either through a change in the cost of sanction ε or the conservatism of the lower court L . The effect of changing these latter parameters is to shift the limit to compliance (\bar{c}_L) in the equality in equation (9).

Corollary 3 *In response to an increase in the lower court’s willingness to comply—either through an increase in the cost of sanction ε or its conservatism L —a compliance-maximizing institutional designer should **increase** the incentive to dissent by:*

- *lowering the cost of dissent d ;¹⁷*
- *decreasing the whistleblower’s share of the sanction α ;*
- *choosing a more conservative whistleblower.*

The corollary states that when the lower court becomes *more* willing to comply, a compliance-maximizing institutional designer should adjust the whistleblower’s parameters to *further encourage dissent*. This is counterintuitive: one might expect that as the lower court’s propensity to comply increases, the need for dissent to inform the higher court of noncompliance would *decrease*.

¹⁷Again, unless it were already 0 in which case she would leave it there.

The reason for this surprising result is that the institution of dissent plays two inter-related, but distinct, roles as a tool to increase compliance. First, it *informs* the higher court that noncompliance has occurred. In this informational role, dissent is a substitute for direct mechanisms of control like increasing sanctions or ideological alignment with the higher court, or both. Second, because the higher court may take a costly action following a dissent, dissent is a *threat*. And, as a threat, dissent can increase compliance even if it is not carried out. But there is a limit to its ability to do so—specifically, the limit to compliance \bar{c}_L derived in Proposition 1. The effect of increasing direct mechanisms of control such as reversal sanctions ε and the lower court’s conservatism L is *both* to increase what compliance would be absent a whistleblower (i.e. \underline{c}_L) and to increase the limit to dissent with a whistleblower (i.e. \bar{c}_L). In other words, increasing direct mechanisms of control increases both compliance *and* the effectiveness of dissent as a threat for inducing even more compliance. The compliance-maximizing institutional response is therefore to have more dissent.

Combined Proof of Proposition 3 and Corollaries 2 and 3 Maximum feasible compliance in equilibrium is \bar{c}_L , and which occurs i.f.f. $c_W^* = \bar{c}_L$. It is straightforward to verify from Proposition 2 that $c_L^* = c_W^* = \bar{c}_L$ is an equilibrium if and only if the equality in Proposition 3 holds.

To prove the institutional design comparative statics in Corollaries 2 and 3, note first that \bar{c}_L is a function of L and ε that is implicitly defined as $\bar{c}_L(\varepsilon, L) = L + \frac{\varepsilon(H - \bar{c}_L(L, \varepsilon))}{k - (H - \bar{c}_L(L, \varepsilon))}$. It is easy to verify that $\bar{c}_L(\varepsilon, L)$ is strictly increasing in L and ε . We now prove comparative statics on the compliance-maximizing choice of W , which we denote $\bar{W}(d, \alpha, \varepsilon, L)$. From the proposition, this quantity is defined as

$$\bar{W}(d, \alpha, \varepsilon, L) - \alpha\varepsilon = \bar{c}_L(\varepsilon, L) + \frac{d\bar{k}}{\phi_d(\bar{c}_L(\varepsilon, L), \bar{c}_L(\varepsilon, L)) - \phi_{nd}(\bar{c}_L(\varepsilon, L), \bar{c}_L(\varepsilon, L))}.$$

The properties in Corollary 2 clearly follow from the fact that l.h.s. is increasing in W and decreasing in α , and the r.h.s. is increasing in d . The properties in Corollary 3 follow from the fact that the r.h.s. is increasing in \bar{c}_L (which in turn follows from Lemma 9) and that

$\bar{c}_L(\varepsilon, L)$ is strictly increasing in L and ε . Nearly identical steps prove the properties for $\bar{\alpha}(W, d, \varepsilon, L)$, the compliance-maximizing share of the sanction.

A slight wrinkle arises for the compliance-maximizing cost of dissent, since d cannot go below 0. First we define $\bar{d}(W, \alpha, \varepsilon, L)$ to be the dissent cost satisfying

$$W - \alpha\varepsilon = \bar{c}_L(\varepsilon, L) + \frac{\bar{d}(W, \alpha, \varepsilon, L) \cdot \bar{k}}{\phi_d(\bar{c}_L(\varepsilon, L), \bar{c}_L(\varepsilon, L)) - \phi_{nd}(\bar{c}_L(\varepsilon, L), \bar{c}_L(\varepsilon, L))}.$$

$\bar{d}(W, \alpha, \varepsilon, L)$ has straightforward comparative statics like the previous implicit characterizations. Moreover, when $\bar{d}(W, \alpha, \varepsilon, L) > 0$ it is the compliance-maximizing dissent cost.

Next we argue that when $\bar{d}(W, \alpha, \varepsilon, L) < 0$, the compliance-maximizing dissent cost is 0. To see this, recall from the proof of Lemma 6 that the compliance-maximizing equilibrium whistleblowing cutpoint $\tilde{c}_W(W, d, \alpha)$ is decreasing in d , and the lower court's best-response $c_L^*(c_W)$ is increasing in c_W when $c_W < \bar{c}_L$. Hence, to show that increasing d above 0 decreases maximum equilibrium compliance only requires showing that $\tilde{c}_W(W, 0, \alpha) < \bar{c}_L$. This follows from the observations that 1) $\tilde{c}_W(W, 0, \alpha) = W - \alpha\varepsilon$, and 2) $\bar{d}(W, \alpha, \varepsilon, L) < 0$ implies that $W - \alpha\varepsilon < \bar{c}_L(\varepsilon, L)$.

Finally, since the compliance-maximizing dissent cost is $\max\{0, \bar{d}(W, \alpha, \varepsilon, L)\}$, and the function $\bar{d}(W, \alpha, \varepsilon, L)$ satisfies the desired monotone comparative statics, $\max\{0, \bar{d}(W, \alpha, \varepsilon, L)\}$ must also (weakly) satisfy them. ■