Uniformity of Interpretation in a Hierarchical Court

Joshua B. Fischman*

January 14, 2006

Abstract

This paper presents a model of legal interpretation in a hierarchical court. Using a two-level court in which judges have spatial preferences over doctrine, the model examines how appeals, panels, and other structural features of the court affect the incentives of judges and promote uniform interpretation of the laws. The threat of appeal has a moderating influence on judges in the lower court. When the cost of appeal is low, this effect will be stronger, but the lower court will also have less influence on the final decision. Hence, under many conditions, overall uniformity will be maximized at an intermediate cost of review. Factors that may increase the predictability of rulings on the higher court, such as panel size, may weaken the incentives toward moderation on the lower court.

Most systems of adjudication feature a hierarchical organization and a process for appeals. While appeal is usually characterized as a procedural safeguard for the benefit of litigants, it clearly also serves as a monitoring device to enforce restraint among judges in the lower courts. This is especially important in an independent judiciary, where there are few constraints on judges' exercise of authority. This article develops a model of judicial interpretation to examine how the appeals process influences judges' incentives. When judges have different ideologies and preferences over methods of interpretation, the model shows that appeals can have the effect of promoting uniformity and ideological moderation

*Department of Economics, Massachusetts Institute of Technology, fischman@mit.edu. I would especially like to thank Glenn Ellison as well as Bengt Holmström, Jim Snyder, Jean Tirole, Mathias Dewatripont, Matthew Headrick, J.J. Prescott, Jackie Chou, and members of lunch and seminar groups at MIT for their helpful comments.

in the interpretation of the law. The model explores how changes in appellate structure, such as panel size and frequency of review, can have an effect on the uniformity of legal interpretation.

Viewing judges as having preferences on an ideological spectrum, the model shows how the appeals process will lead to rulings that are clustered in the middle of the spectrum. This is not only because of "correction" of extreme rulings by an appeals court, but also because the prospect of appellate review leads lower courts to moderate their rulings. Appeals have the effect of depoliticizing judicial rulings by mitigating the influence of individual judges' ideologies in the outcome of cases. Thus the outcome of a case will be less dependent on the identity of the judge hearing the case.

This is important for several reasons. First, basic conceptions of fairness and rule of law require similar outcomes for similar litigants. If a judge's idiosyncratic preferences were a major factor in each decision, the application of the law would be viewed as arbitrary and illegitimate. Second, uniformity of interpretation promotes predictability of the law, allowing people and firms to make better plans in the face of legal uncertainty. Third, when there is uncertainty regarding how a law will be applied, compliance will be uneven, and the law will be less effective at achieving its stated goals. Finally, uniformity reduces litigation: parties will be less likely to have conflicts and more likely to settle if they have similar expectations about how the law will be applied.

This is not to suggest that absolute uniformity is possible, or even desirable. Variability is a necessary consequence of judicial discretion. Disagreements among judges leads courts to reexamine and modernize precedent and facilitates the evolution of the law. Also, uncertain laws could potentially lead to better compliance among risk-averse agents. The point to be made here is simply that idiosyncratic interpretation by judges can be harmful to the rule of law, and that the organization of courts has an impact on how ideologically heterogeneous judges harmonize their rulings.

Previous models of hierarchical courts have either focused on the role of error correction, or examined how higher courts enforce compliance with doctrine in the lower courts. The error correction models, e.g., Shavell (1995), Daughety & Reinganum (2000), Cameron & Kornhauser (2005), view judges as being part of a "team" who share the same objective: maximizing the number of correctly decided cases. While these models explain one essential purpose of the appellate system, they are less applicable to the interpretive, or "lawmaking" function of courts. In areas where the law is indeterminate, determining what the law means is an exercise in judicial discretion. In this context, ideological differences among judges are more significant, and the assumption that judges share the same objective is weakest.

To analyze the lawmaking context, this article employs the "political" model of judging[1], in which rulings are viewed as ideological decisions that can be mapped into a one-dimensional "policy space." Using a two-level hierarchy in which there is uncertainty about the higher courts' preferences, the model shows how lower court judges will strategically shift their rulings toward the center of the spectrum. Previous models of political judging in a hierarchy have investigated various ways in which higher courts monitor lower courts in order to enforce doctrinal compliance. (McNollgast 1995, Cameron, Segal & Songer 2000, Spitzer & Talley 2000, Mialon Rubin & Schrag 2004) The contribution of this article is in examining how the interaction between courts promotes uniformity in the interpretation of the law. This has been an important concern in the legal literature, especially in the context of analyzing proposed structural changes in the courts.[2] By developing a formal model with testable implications, this article seeks to bring the insights of economics and positive political theory to the analysis of court structure and judicial interpretation.

Although the article focuses on courts, the insights of the model can be applied to any situation involving ideological decision-making and serial review. Thus it could also apply to relations between administrative agencies and courts, agencies and legislatures, or civil servants and political appointees.

We summarize the main conclusions of the model here. First, judges on the lower court will shift their rulings toward the center of the distribution of interpretations. This shift represents a balance between two interests: conformity with the judge's own preferences and likelihood of surviving review. Thus the appeals process can increase the predictability of interpretation, even when appeal occurs infrequently. Second, more frequent review will create stronger incentives for the lower court to rule moderately, but will also result in more cases being decided by the higher court, which is unconstrained.. Thus it will always be optimal to limit appeals to some degree. Third, increasing the consistency of rulings in the

[1] This is also sometimes referred to as the "attitudinal" model. (Segal and Spaeth 2002)

[2] Examples include the debate over splitting the Ninth Circuit Court of Appeals (e.g., Hug 2000, O'Scannlain 1999), changes to the en banc review process (Banks 1997), procedural reforms for immigration appeals (ABA Commission on Immigration Policy 2003), and a now-shelved proposal to establish an Intercircuit Panel to resolve conflicts of law between circuits (Ginsburg and Huber 1987), have focused on the how these plans would affect consistency and coherence in the law.

higher court (for example, by using large panels) may increase the consistency of the lower court, but this effect is nonmonotonic: too much consistency in the higher court will have the opposite effect. When the higher court is too predictable, the lower court judges can "game" the appeals process: they will know more precisely how much they can adhere to their preferences while still avoiding appeal.

Section I provides the setup of the model. Section II provides some general results and shows that under general assumptions, all lower court judges will avoid rulings that are outside a bounded set within the spectrum of interpretations. In section III, we make parametric assumptions regarding judges' preferences, and derive some implications of the model. In particular, we can show that a very high frequency of appeal or a high court that is too predictable can be suboptimal for the uniformity of final rulings. Finally, section IV also provides some numerical calculations and graphs to illustrate the effects of court structure and the interplay between appeals and panels.

I. Description of the Model

This model focuses on the process of legal interpretation. Hence, we may take issues of fact as having been predetermined by the court. We will assume that it is a dominant strategy for the losing party to appeal, and that appeals are at the discretion of the higher court.[3]

Following other spatial models of hierarchical courts (McNollgast 1995, Cameron, Segal & Songer 2000, Spitzer & Talley 2000, Segal & Spaeth 2002) and legislature-agency interactions (e.g. Ferejohn & Shipan 1990, Eskridge & Ferejohn 1992), we use a one-dimensional spatial model of judging. Each ruling is represented as a number on the real line. We can think of this spectrum as representing the range of plausible interpretations of the law in question, with $+\infty$ and $-\infty$ representing the extremes. For example, the spectrum could represent "liberal" vs. "conservative" preferences, or "strict" vs. "loose" construction of laws. Although courts are typically hierarchical with multiple levels, we consider a two-level court for simplicity. The model assumes that judges' preferences are based strictly on

[3] Although these assumptions are not formally true in every court system, they are still reasonable in situations where the costs of appeal will be small compared to the amounts of money and the importance of the legal issues at stake. Also, even when appeals are automatic, the higher court may only provide perfunctory review in cases where there is no disagreement with the lower court. For example, in the federal system, circuit courts do not have discretion to deny appeals, but they may dispose of cases in unpublished, non-precedential opinions. In the federal circuit courts, an overwhelming majority of cases are in fact decided in unpublished opinions. (Merritt and Brudney 2001)

doctrine – how the law is interpreted – and not on how this doctrine impacts the particular litigants in the case.

The model does not explicitly account for how precedent influence judges' decisions . Since we are only considering cases where the law is indeterminate, we may assume that no precedent is dispositive. If there are multiple precedents that are relevant, then judges' ideal points may reflect the amount of weight that they place on each of these precedents. For example, a judge with an ideal point on the "liberal" side of the spectrum might place more weight on a "liberal" precedent than a judge on the "conservative" side of the spectrum. Also, the assumption that judges' utility is based on the about the final ruling in a case (and not merely which side wins) implies that judges expect that their rulings will have precedential value in subsequent cases.

The model assumes that judges are concerned only with the final disposition of a case, and derive no utility from "posturing." A judge with preferences outside the mainstream would therefore prefer to moderate his opinions to reduce the risk of being overruled by an appellate court, if he expects that the appellate court would deviate even further from his own preferences.

We consider two courts, a lower court consisting of a single judge, and a higher court consisting of a single judge or a panel of judges. In the first stage, the lower court judge issues a ruling $x \in \mathbb{R}$, representing her interpretation of the law. In the second stage, the higher court decides whether to review the ruling; if it does, it issues a new ruling y. If the higher court reviews the lower court's ruling, it will incur an effort cost $e > 0$, and the lower court will incur a disutility $d \geq 0$. This disutility represents a reputational cost to the lower court judge from being overruled.

We assume that the lower court judge's utility function is

$$U_l = \begin{cases} -\left(q_l - x\right)^2, \text{ if it is not overruled} \\ -\left(q_l - y\right)^2 - d, \text{ if it is overruled} \end{cases}$$

where q_l is the lower court judge's ideal point. Similarly, the higher court's utility function is

$$U_h = \begin{cases} -\left(q_h - y\right)^2 - e, \text{ if it overrules the lower court} \\ -\left(q_h - x\right)^2, \text{ if it does not overrule the lower court} \end{cases}$$

where q_h, the higher court's ideal point, is unknown to the lower court judge. We can view this as reflecting random assignment of judges to panels, so that the identity of the appellate judges is unknown to the trial judge, or as uncertainty about the higher court's preferences on this question of law. We model q_h as a random variable with density function f^h and cumulative distribution function F^h, where f^h is continuous, symmetric, and unimodal with mean μ_h. Note that because of symmetry, μ_h will also be the median and the mode of the distribution of q_h. We assume that both distributions have full support over the spectrum of interpretations.

II. General Results

Our first two results explain the basic effect illustrated by our model: that judges on the lower court will strategically shift their rulings toward the center of the distribution, in order to balance their own preferences with the likelihood of appeal.

Our first theorem establishes a simple rule for determining when the higher court will hear an appeal, and characterizes the lower court's ruling implicitly as a function of g.

Theorem 1

If $x \in [q_h - c, q_h + c]$, where $c = \sqrt{e}$, the higher court judge will decline to review the case. If $x \notin [q_h - c, q_h + c]$, the higher court judge will review the case and issue a ruling $y = q_h$. The lower court judge will issue a ruling x that satifies

$$x = g^{-1}(q_l)$$

where

$$g(z) = z + \frac{1}{2} \frac{\left(c^2 + d\right) \left[f^h\left(z + c\right) - f^h\left(z - c\right)\right]}{c\left[f^h\left(z + c\right) + f^h\left(z - c\right)\right] - \left[F^h\left(z + c\right) - F^h\left(z - c\right)\right]}$$

Proof. See Appendix. ■

We can think of c as the amount of latitude that the higher court will accept in the lower court's interpretation. Remember that q_h is unknown to the lower court, so that the lower court knows the size, but not the position, of the interval of permissible rulings.

Although theorem 1 only defines x implicitly as a function of q_l, we can use it to understand the shape of the lower court's choice function.

Theorem 2

The lower court's ruling x will always be between its own ideal point q_l and the center of the appeals court's distribution μ_h. If $q_l = \mu_h$, then $x = \mu_h$.

Proof. See Appendix. ■

This result follows from the fact that the fractional part of $g(x)$ will be strictly positive for $x > \mu_h$ and strictly negative for $x < \mu_h$. Thus when $x > \mu_h$, we have $q_l = g(x) > x > \mu_h$. Similarly, when $x < \mu_h$, we have $q_l = g(x) < x < \mu_h$.

This theorem demonstrates one of the basic results of the model: that judges will shift their rulings from their own preference points toward μ_h, the center of the appeals court's distribution. This shift is motivated by the tradeoff between issuing a ruling close to the judge's preference point and reducing the chance of being overruled by ruling close to the center. Since g is continuous, and $g(\mu_h) = \mu_h$, the amount of shift will be small when q_l is close to μ_h.

Even though the lower court judges' ideal points are distributed over the entire real line, the incentives created by the appeals process will limit their choice of rulings to a bounded range. The intuition for this is that if the lower court judge chooses a ruling that is extreme, it will have a very high probability of being overruled. Outside the interval (x_0, x_1), this effect will dominate the benefit to the lower court judge of choosing a ruling close to her ideal point. Instead, the judge would prefer a ruling closer to μ_h that has a greater chance of surviving review. We state this result formally in the following theorem:

Theorem 3

There exist x_0 and x_1 such that the lower court's ruling will always be bounded by the interval (x_0, x_1), regardless of q_l.

Proof. See Appendix. ■

The proof relies on the fact that the denominator of the fractional part of g will be negative at μ_h, but will be positive in some range on either side of μ_h. Hence g will have asymptotes on both sides of μ_h. Even though the lower court judges' ideal points are distributed over the entire real line, the incentives created by the appeals process will limit their choice of rulings to a bounded range.

Theorem 3 shows that there is a subset of interpretations – those outside the interval (x_0, x_1) – that will *never* be chosen by a lower court judge, even though they are preferred by a non-trivial subset of lower court judges and have a positive chance of being upheld.

These theorems show how the appeals process leads to more uniform interpretation of the law by the lower court. First, instead of ruling at his own ideal point, the lower court judge will choose a ruling that is closer to μ_h; this shift will typically bring judges with different preferences closer together. Second, the set of possible rulings issued by the lower court will be bounded. This will eliminate the possibility of the most extreme interpretations being issued by the lower court. Note that in our model, there are no constraints on the higher court, so that its rulings will be unbounded if f^h has full support on the real numbers. Nevertheless, the likelihood of the most extreme rulings will be reduced significantly.

Although this result seems counterintuitive – the lower court is always centrist, while the appeals court may be extreme – it captures the fact that major changes in doctrine usually issue from the highest court. Lower courts do not have the authority to dramatically alter the interpretation of the law, as higher courts can.

Theorems 2 and 3 also show how judicial restraint derives from the incentives facing the lower court. The lower court judge's interest in influencing the law, and its awareness that it may be overruled if it strays too far from the center, provide strong incentives to suppress its personal views. Note that both theorems hold irrespective of d, even if there is no additional disuitility from being overruled.

Theorem 2 shows how the appeals process causes lower court rulings to be clustered around μ_h. The use of panels will also cause rulings from the higher court to be more tightly clustered around μ_h. Thus both appeals and panels lead rulings to be closer to the ideal point of the median appellate judge. This focal point has several appealing consequences. For example, if we assume that there exists an optimal interpretation, which each judge observes with error, then the median will closely approximate the optimum. One the other hand, if we view interpretations of the law as political choices, then the median represents the most democratic outcome. Finally, the view of the median judge has the benefit of legitimacy, in that represents a consensus decision among judges.

The following theorem states the density function of the final outcome y. We will use this in Section IV of the paper to graph some results of the model when both courts' ideal points are normally distributed.

Theorem 4

Let F^h and F^l denote the cumulative density functions of q_h and q_l, respectively, and f^h and f^l denote the density functions. If $g(x)$ is an increasing function, then the density *function f^y of y satisfies*

$$\begin{aligned} f^y(y) &= \left[F^h(y+c) - F^h(y-c)\right] f^l(g(y))\, g'(y) \\ &\quad + \left[1 - F^l(g(y+c)) + F^l(g(y-c))\right] f^h(y) \end{aligned}$$

Proof. See Appendix ■

We provide an outline of the proof. There are two ways of reaching a final ruling of y: a lower court ruling of y that is not overruled, or a different lower court ruling, followed by a higher court ruling of y.

In the first case, we have $q_l = g(y)$. The probability that the higher court will not review is $\Pr(|q_h - y| \leq c) = \left[F^h(y+c) - F^h(y-c)\right]$. This represents the first term in the right-hand side of theorem 4.

In the second case, we have $q_l = g(x)$, $q_h = y$. The probability that the higher court will review is

$$\begin{aligned} \Pr(|x-y| > c) &= \Pr(y - c < x < y + c) \\ &= \Pr(g(y-c) < q_l < g(y+c)) \\ &= 1 - F^l(g(y+c)) + F^l(g(y-c)) \end{aligned}$$

The contribution from the second way of reaching a ruling of y represents the second term in theorem 4.

III. Implications of the Model

In this section, we will explore some of the implications of the model. By making parametric assumptions – normally distributed preferences on both levels of the court – we can examine the impact of changes in the cost of review and the predictability of the higher court. In part (a), we show that the cost of review will have nonmonotonic effects on the variance of higher court rulings. This results from two competing effects: lower court judges will be more restrained when the cost of review is low, but since they are more likely to be overruled, their decisions will have less impact on the final ruling. Thus the variance of final rulings will always be minimized at $c > 0$.

In part (b) we prove another nonmonotonicity theorem: that the predictability of the higher court σ will have nonmonotonic effects on the range of lower court decisions. When the higher court's rulings are highly variable, the lower court will have less incentive to be moderate, since the likelihood of being overruled will decrease less sharply toward the center. On the other hand, if the higher court is very predictable, the lower court can more easily "game" the higher court. Since there will be a fairly well-defined "safe" region, in which the likelihood of review is low, the lower court can shift toward the center only as much as necessary to avoid review.

These results are presented in terms of parameters c and σ, representing the latitude granted to the lower court and the predictability of the higher court, respectively. Since $c = \sqrt{e}$, where e is the equilibrium effort cost of review, we can use these results to inform our understanding of the effects of structural changes in the court. Changes that reduce the equilibrium cost of review, such as appointing additional judges, or reducing the total caseload, could be modeled as decreasing c. Increasing the size of appellate panels or employing a more rigorous screening process for judicial appointments could be modeled as reducing σ.

A. Cost of Review

We assume that

$$q_l \sim N(0,1)$$

and

$$q_h \sim N(0,\sigma^2)$$

Implicit in these distributional assumptions is that the preferences of judges on both courts have equal means. We assume that there is no ideological tension between the courts in order to focus on the effects of court structure on uniformity of interpretation.

Recall that the lower court's ruling is determined by $q_l = g(x)$, where

$$g(x) = x + \frac{1}{2}\frac{(c^2+d)\left[f^h(x+c) - f^h(x-c)\right]}{c\left[f^h(x+c)+f^h(x-c)\right] - \left[F^h(x+c) - F^h(x-c)\right]}.$$

where $f^h(x) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)$. Recall from theorem 3 that the bounds of x (the possible rulings of the lower court) are determined by the denominator of the above expression. By symmetry, we may denote the lower and upper bounds of x as $-\lambda$ and λ. In this section, λ will be endogenously determined as a function of c and σ.

The following theorem shows how the cost of appeal affects the range of rulings issued by the lower court.

Theorem 5

The range of lower court rulings $(-\lambda, \lambda)$, where the cost of review is c and the variance of the higher court rulings is σ, has the following properties with respect to c:

a) $\lambda(c, \sigma)$ *is finite for all* c, σ

b) $\lim_{c \to 0} \lambda(c, \sigma) = \sigma$.

c) $\lambda(c, \sigma) \sim c$ *as* $c \to \infty$.

d) λ *is strictly increasing in* c.

Proof. See Appendix. ■

Note that Part (a) is a direct result of theorem 3. Part (b) shows that the range will contract to $[-\sigma, \sigma]$ as $c \to 0$, but even when the probability of review approaches 1, the range of rulings will not contract completely.

Part (c) holds because as c gets large relative to σ, any lower court judge with $|q_l| < c$ will rule very close to his ideal point, since the probability of review will be very small. Judges with $|q_l|$ close to c will shift just enough to reduce the probability of review to be very small. Hence the outer bound λ will be close to c in magnitude.

Part (d) is an intuitive result: the bounds of possible rulings will expand when the cost of review increases.

Taken together, these results show the relationship between c and the bounds of the lower court's rulings.

Theorem 5 showed a mononoticity result on lower court rulings: that the lower court becomes more predictable when c decreases. However, if appeal is too frequent, then the lower court will only have a small impact on the final ruling, since its decision will usually be reviewed. Thus a low value of c provides strong incentives for the lower court to be rule moderately, but the predictability of the lower court will not matter for the final result. The following theorem formalizes the above intuition, showing that the variance of the final ruling is minimized at a positive cost of review.

Theorem 6

The variance of the final ruling y will be decreasing in c as $c \to 0$, and will therefore be minimized at some $c > 0$.

Proof. See Appendix. ∎

B. Uncertainty

In this section, we will study the effect of σ on the rulings of the lower court. Since σ is the standard deviation of higher court rulings, we can use it to model structural changes in the court, such as panel size, changes in the process for appointing judges, or the impact of external constraints such as legislative override or judicial elections. As the following theorem the effect of σ on the lower court rulings is nonmonotonic: the bounds of rulings will be decreasing in σ for small σ and increasing in σ for large σ.

Theorem 7

The range of lower court rulings $(-\lambda, \lambda)$ has the following properties with respect to σ:

a) $\lim_{\sigma \to 0} \lambda(c, \sigma) = c$.

b) λ is nonmonotonic with respect to σ: it is decreasing for small σ and increasing for large σ.

c) For any fixed c, λ reaches a global minimum at some $\sigma > 0$.

Proof. See Appendix. ∎

Part (a) examines the bounds when $\sigma \to 0$, i.e, when the higher court's decision approaches certainty. In this case, any ruling x satisfying $|x| \leq c$ will not be reviewed by the higher court. Therefore the lower court judge will rule sincerely if $|q_l| \leq c$. Any judge with $q_l > c$ will choose $x = c$, and any judge with $q_l < -c$ will choose $x = -c$. Thus the range of lower court rulings will be $[-c, c]$.

Part (b) shows that the range of the lower court's rulings will *decrease* in σ for low values of σ and *increase* for high values of σ. This is a somewhat counterintuitive result: for low values of σ, as the higher court rulings become less uniform, the lower court rulings become more uniform. The intuition for this is as follows: when $\sigma = 0$, the higher court's position is known, so the lower court will know exactly how much to moderate its opinion, if necessary, in order to avoid appeal. In particular, if the lower court's preference point q_l satisfies $|q_l| > c$, then the lower court will choose $x = \pm c$. If we increase σ very slightly, then there is a very small amount of uncertainty about the higher court's preferences. If the lower court now chose $x = \pm c$, there would be a $\frac{1}{2}$ probability of review. However, by moving slightly toward the center, the probability of review drops dramatically; at $x = \pm(c - \sigma)$, the probability of review goes to 0. The decreased likelihood of review results in a first-order gain, while the shift from the preference point results in a second-order loss. Thus the lower court will shift slightly toward the center in such cases.

Part (c) is an immediate consequence of part (b).

This means that reducing the uncertainty of the appellate judges' decisions will increase the outer bound of the lower court's ruling when the level of uncertainty is already very low.

C. Aversion to Reversal

Theorem 8

The range of lower court rulings $(-\lambda, \lambda)$ is independent of d the disutility from being overruled, but the variance of lower court rulings is strictly decreasing in d.

Proof. See Appendix. ■

The result shown in theorem 8 is unsurprising: that lower court judges will rule more moderately when they are averse to being overruled by the higher court. What is most significant is that the model does not require any individual disutility from reversal to show that appeals promote moderation in the lower courts. The desire to influence the outcome of cases, coupled with strategic anticipation of the appellate court's ruling, is sufficient to ensure moderation in the lower court.

Although it is frequently believed that judges do not like to be overturned, i.e., $d > 0$, this assumption has been challenged. (Klein and Hume 2003) Thus, although incorporating reversal aversion strengthens the predictions of the model, it is not necessary for any of the results in this section.

IV. Effects of Court Structure on Interpretation

In this section, we use some of the results from Section III to explore the effects of court structure on interpretation. As in Section III, we assume that we have a two-level court system, and we normalize the variance of the lower court judges' ideal points to 1. Let σ^2 be the variance of the higher court, where typically $\sigma^2 < 1$. Finally, we assume that $d = 0$.

Appeals courts usually consist of a panel of judges. Formally, panels rule by majority vote, but the deliberative process and collegial decision-making among the judges may lead to a more nuanced process of preference aggregation. (Sunstein, Schkade, and Ellman 2004) In a purely independent voting model, the higher court ruling would be the median ideal point among the judges in the panel; under a joint utility maximization model, the ruling would be the mean.

There are other factors that would influence σ^2. The appointment and confirmation process could lead to greater scrutiny of potential judges, and hence more moderation, or a highly politicized process could have the opposite effect. Influences from outside the judiciary – such as the threat of legislative override, greater media scrutiny, and in some case, judicial elections – may have a stronger influence on the appeals court. Without attempting to model each of these effects explicitly, we may simply observe that the predictability of the higher court should be increasing in panel size, and that the variance of a court with panel size n should decrease proportionately by a factor of order n.

For simplicity, we assume in the following discussion that $\sigma^2 = \frac{1}{n}$. This would occur, for example, if all judges' preferences on both courts are drawn from the same distribution, and the higher court chooses a ruling that maximizes the sum of utilities of all judges on the panel.

First, we analyze the simplest model: a higher court with a single judge, so that $\sigma^2 = 1$. At $c = 0$, the lower court has no latitude and appeal is costless, so that the outcome will always coincide with the higher judge's ideal point. At $c = \infty$, appeal will never occur, and the trial judge will always rule at her ideal point. Thus, in either of these cases, the outcome will be normally distributed with variance 1. For $0 < c < \infty$, the variance will be strictly less than 1, as shown in Figure 1. Notice that the probability of review decreases in c, as expected.

Figure 1 shows that the variance in this case is minimized at $c \approx 2.3$, which corresponds to a likelihood of review of around 10%. At this level, the variance of the final outcome is 0.52. When both courts have the same variance, uniformity is optimized at a relatively low likelihood of review.

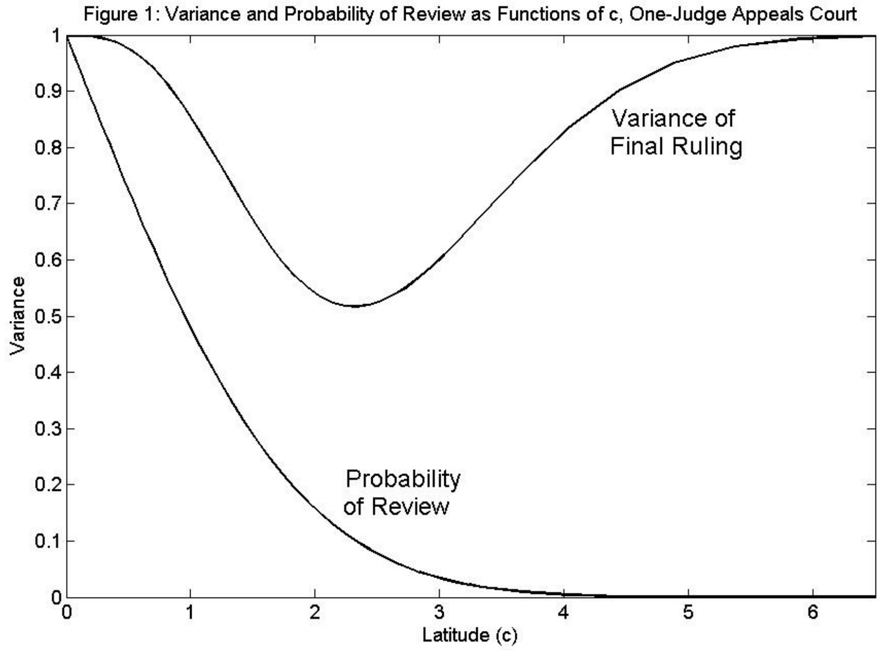

Figure 1:

Next, we consider the effect of using 3-judge panels in the higher court, so that $\sigma^2 = \frac{1}{3}$. Figure 2 shows the impact of c on the variance of the outcome and the probability of review. Here, the minimum variance is approximately 0.17 at $c \approx 1.1$; the probability of review is approximately 0.34. At this level, the variance is reduced by a factor of almost 6. Since the probability of review is 0.34, and each appeal requires 3 judges, the court system will require the same level of resources ("judge-hours") on the higher court as on the lower court.

The degree of latitude is not directly controlled; it is determined in equilibrium by the marginal effort cost of the higher court judges. Suppose, for example, that there are equal numbers of judges on both courts, and that each case on the higher court requires a panel of three judges. In equilibrium, if each higher court judge is participating in as many cases as each lower court judge, then the probability of review must be $\frac{1}{3}$. This would mean that $c \approx 1.1$, which is close to the optimum.

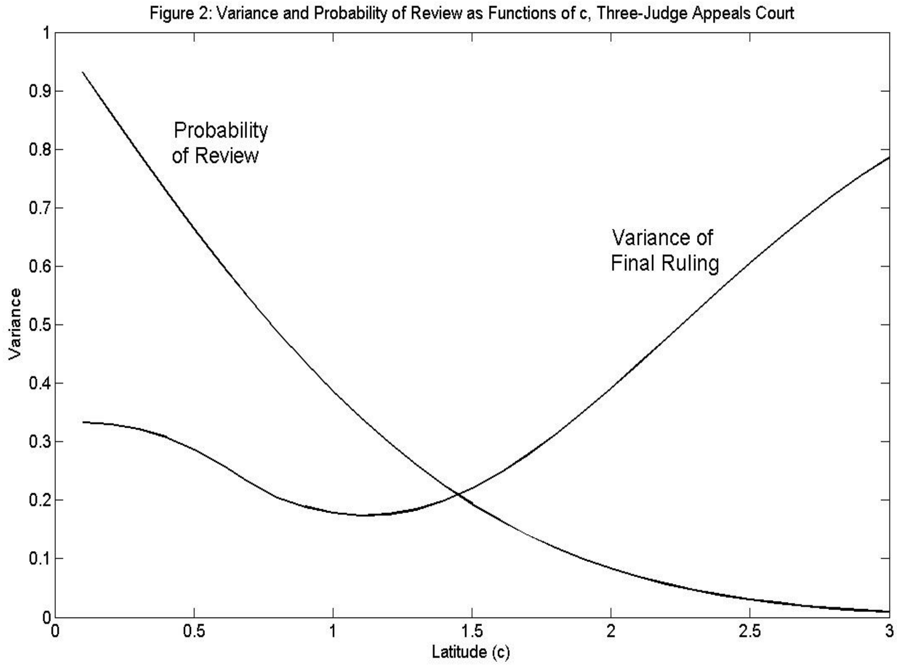

Figure 2:

Figure 3 illustrates how lower variances can be achieved with larger panels and a greater mass of judges. It compares the variance against the measure of judges for panels of one, three, and five judges. The x-axis in Figure 3 measures "Ratio of Higher/Lower Court Workload," which is scaled under the assumption that each judge's participation in an appeal requires the same amount of effort as a lower court judge requires to decide a case.

Note that when $c = \infty$, the variance will be equal one in all cases, since there will never be appeal. This corresponds to the point in the upper-left corner of the graph. In the case of a one-judge higher court, the variance goes to one as $c \to 0$; this corresponds to the scenario where every case is decided by the appeals court (when the workload ratio is 1).

The superimposed variances for different panel sizes shows that larger reductions in variance are possible with larger panels, but that more appellate judges are necessary to achieve these reductions. Thus when few appellate judges are used, smaller panels with more frequent review will yield more uniformity than larger panels with less frequent review. For example, if the ratio of higher court to lower court judges is 0.5, Figure 3 shows that 3-judge panels achieve the greatest degree of uniformity. When many judges are used, it is possible to achieve large reductions in variance when there are both large panels and

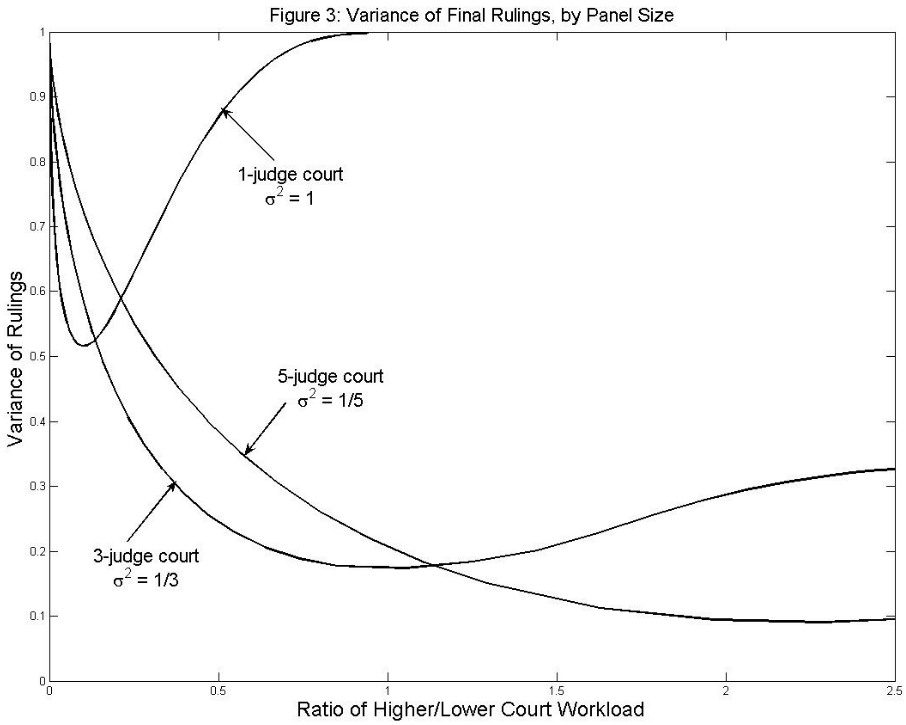

Figure 3:

frequent review.

It is worth noting that, at least in theory, significant reductions in variance may occur even when the likelihood of review is very small. For example, the U.S. Supreme Court agrees to hear fewer than 3% of cases that are petitioned for review.(Epstein, Segal, Spaeth, and Walker 1996) According to the model, this could still have a significant impact on judges in the appeals courts.

Figure 3 presents a very simplified graph illustrating how appeals may increase the uniformity of final rulings. The reduction in variance would be greater when judges are averse to being overruled ($d > 0$). The graph also does not consider the effect of a multiple-tiered hierarchy, which could achieve even greater reductions in variance. Nevertheless, it provides a sense of how choice of court structure can affect how judges interpret the law.

V. Conclusion

Formal modeling in this context provides several insights that would be less obvious from casual observation. When there is uncertainty about the ideologies of judges on both tiers of the court, then the strategic interaction between the lower court and the higher court can result in greater consistency of interpretation that either court could achieve independently. Thus, even when the higher court is much more consistent, it is still optimal to limit review to some degree.

In the legal literature, it is often assumed that more frequent review enhances predictability in the interpretation of the law by providing closer monitoring of lower court judges and resolving conflicts arising from different cases. This assumption neglects the fact that higher court itself introduces some uncertainty. For example, while the Supreme Court has been criticized for not reviewing enough cases, commentators have also complained of doctrinal incoherence in areas of the law in which the Supreme Court has been active. (Hellman 1996).

Similarly, questions of panel size are typically framed as a trade-off between consistency and the use of judicial resources. As the model shows, however, using panels to increase consistency in appellate courts may be counterproductive if it weakens incentives for moderation in the lower court. Whether this effect is observable in practice is a question for future empirical research.

There are several theoretical questions arising from this model that could be explored in future research. The model could be extended to consider multiple-tiered hierarchies, which could potentially provide strong incentives for restraint in the lower court, while minimizing concentration of authority in the highest court. A model that endogenizes the role of precedent could help explain how appeals judges select cases for review, and also explore how the structure of courts affects the evolution of precedent.

VI. Appendix

Proof of Theorem 1:

First, note that for a given ruling x, the higher court judge has utility $-(q_h - x)^2$ if he does not review the case, and utility $-(q_h - y)^2 - e$ if he does review. Clearly, this latter term is minimized at $y = q_h$, so the judge will review the case if $-e > -(q_h - x)^2$. Hence the ruling will be reviewed if $|x - q_h| > \sqrt{e} = c$.

Now for a given ruling x, the lower court judge's utility will be

$$U_l = \begin{cases} -\left(q_l - q_h\right)^2, \text{ if the case is reviewed} \\ -\left(q_l - x\right)^2, \text{ if the case is not reviewed} \end{cases}$$

Thus,

$$\begin{aligned} EU_l &= \int_{-\infty}^{x-c} -\left(q_h - q_l\right)^2 f^h\left(q_h\right) dq_h + \int_{x-c}^{x+c} -\left(x - q_l\right)^2 f^h\left(q_h\right) dq_h \\ &\quad + \int_{x+c}^{\infty} -\left(q_h - q_l\right)^2 f^h\left(q_h\right) dq_h \end{aligned}$$

Applying Leibnitz's Rule,

$$\begin{aligned} \frac{\partial EU_l}{\partial x} &= -\left(x - q_l - c\right)^2 f^h\left(x - c\right) - 2\left(x - q_l\right)\left[F^h\left(x + c\right) - F^h\left(x - c\right)\right] \\ &\quad -\left(x - q_l\right)^2\left[f^h\left(x + c\right) - f^h\left(x - c\right)\right] + \left(x - q_l + c\right)^2 f^h\left(x + c\right) \\ &= 2\left(x - q_l\right)\left[c\left[f^h\left(x + c\right) + f^h\left(x - c\right)\right] - \left[F^h\left(x + c\right) - F^h\left(x - c\right)\right]\right] \\ &\quad + \left(c^2 + d\right)\left[f^h\left(x + c\right) - f^h\left(x - c\right)\right] \end{aligned}$$

Hence the first-order condition yields

$$x = q_l - \frac{1}{2}\frac{\left(c^2 + d\right)\left[f^h\left(x + c\right) - f^h\left(x - c\right)\right]}{c\left[f^h\left(x + c\right) + f^h\left(x - c\right)\right] - \left[F^h\left(x + c\right) - F^h\left(x - c\right)\right]}$$

To simplify our notation, let

$$\begin{aligned} g\left(x\right) &= x + \frac{1}{2}\frac{\left(c^2 + d\right)\left[f^h\left(x + c\right) - f^h\left(x - c\right)\right]}{c\left[f^h\left(x + c\right) + f^h\left(x - c\right)\right] - \left[F^h\left(x + c\right) - F^h\left(x - c\right)\right]} \\ n\left(x\right) &= \frac{1}{2}\left(c^2 + d\right)\left[f^h\left(x + c\right) - f^h\left(x - c\right)\right], \text{ and} \\ d(x) &= c\left[f^h\left(x + c\right) + f^h\left(x - c\right)\right] - \left[F^h\left(x + c\right) - F^h\left(x - c\right)\right]. \end{aligned}$$

so that

$$q_l = g(x) = x + \frac{n(x)}{d(x)}$$

Proof of Theorem 2:

Let $n(x)$ and $d(x)$ denote the numerator and denominator, respectively, of $g(x)$:

$$\begin{aligned} n(x) &= \frac{1}{2}(c^2+d)\left[f^h(x+c)-f^h(x-c)\right], \text{ and} \\ d(x) &= c\left[f^h(x+c)+f^h(x-c)\right]-\left[F^h(x+c)-F^h(x-c)\right] \end{aligned}$$

so that

$$q_l = g(x) = x + \frac{n(x)}{d(x)}$$

Then $g(x)$, $n(x)$, and $d(x)$ satisfy the following properties:

a) $n(x) \geq 0$ for $x \leq \mu_h$

b) $n(x) \leq 0$ for $x \geq \mu_h$

c) $d(\mu_h) < 0$

d) $g(\mu_h) = \mu_h$

First, we will show that $n(x) \geq 0$ for $x \leq \mu_h$:

For $x < \mu_h - c$, the result follows from the fact that f^h is increasing on $(-\infty, \mu_h)$. For $\mu_h - c < x < \mu_h$, note that

$$\begin{aligned} n(x) &= \frac{1}{2}(c^2+d)\left[f^h(x+c)-f^h(x-c)\right] \\ &= \frac{1}{2}(c^2+d)\left[f^h(2\mu_h-x-c)-f^h(x-c)\right] \\ &> 0, \text{ since } x-c < 2\mu_h-x-c < \mu_h. \end{aligned}$$

Similarly, we can show that $n(x) \leq 0$ for $x \geq \mu_h$. This also implies that $n(\mu_h) = 0$.

Now we show that $d(\mu_h) < 0$:

Note that $f^h(\mu_h - c) = f^h(\mu_h + c)$, and f^h is increasing on $(\mu_h - c, \mu_h)$ and decreasing on $(\mu_h, \mu_h + c)$. Thus $f^h(x) \geq f^h(\mu_h - c)$ for all $x \in (\mu_h - c, \mu_h + c)$, with strict inequality in a neighborhood of μ_h. Hence,

$$\begin{aligned} d(\mu_h) &= c\left[f^h\left(\mu_h+c\right)+f^h\left(\mu_h-c\right)\right]-\left[F^h\left(\mu_h+c\right)-F^h\left(\mu_h-c\right)\right] \\ &= c\left[f^h\left(\mu_h-c\right)+f^h\left(\mu_h-c\right)\right]-\int\limits_{\mu_h-c}^{\mu_h+c} f^h(t)dt \\ &= \int\limits_{\mu_h-c}^{\mu_h+c}\left[f^h(\mu_h-c)-f^h(t)\right]dt \\ &< 0. \end{aligned}$$

Also, this implies that $\frac{n(\mu_h)}{d(\mu_h)}=0$ so that $g(\mu_h)=\mu_h$.

Now d must be negative over the range of x. If $d(x_1)=0$ for some value $x_1>\mu_h$, then g will asymptote to $+\infty$ at x_1. Similarly, g will asymptote to $-\infty$ if $d(x_0)=0$ for some value $x_0>\mu_h$. Hence for all x for which $g(x)$ is defined, $d(x)<0$.

Since $n(x)\le 0$ for $x>\mu_h$ and $d(x)<0$, it follows that $g(x)>x$, and therefore, $\mu_h<x<q_l$. Similarly, for $x<\mu_h$, we have $q_l<x<\mu_h$. Therefore the judge chooses a ruling x between q_l and μ_h.

Proof of Theorem 3:

We will show that g has asymptotes on both sides of μ_h. First, note that

$$\begin{aligned} \int\limits_{-\infty}^{\infty} d(x)dx &= \int\limits_{-\infty}^{\infty} c\left[f^h(x+c)+f^h(x-c)\right]dx-\int\limits_{-\infty}^{\infty}\int\limits_{x-c}^{x+c} f^h(t)dtdx \\ &= 2c-2c\int\limits_{-\infty}^{\infty} f^h(t)dt \\ &= 0 \end{aligned}$$

Since d is continuous, there exists some x_0 such that $d(x_0)=0$.

Also, $x_0\neq\mu_h$, since $d(\mu_h)<0$, as shown in theorem 2. If there exist multiple choices for x_0, choose the value that minimizes $|\mu_h-x_0|$. Since f is symmetric about μ_h, it follows that d is symmetric about μ_h, so that $d(x_1)=0$, where $x_1=2\mu_h-x_0$. For simplicity, we can assume that $x_0<\mu_h<x_1$. Now $n(x_1)<0$ and $n(x_0)>0$, so $\lim_{x\to x_0^+}\frac{n(x)}{d(x)}=-\infty$ and

$\lim_{x \to x_1^-} \frac{n(x)}{d(x)} = +\infty$. Thus g has asymptotes at x_0 and x_1 and has a range covering the entire real line on (x_0, x_1). Since $q_l = g(x)$, it follows that $x_0 < x < x_1$. Thus the lower court's choice of ruling will always be bounded by the interval (x_0, x_1).

Proof of Theorem 4:

Consider any value y_0. Then $F^y(y_0) = \Pr(y < y_0)$. Now there are two conditions on q_l and q_h that will achieve $y < y_0$: either $x < y_0$ and the case is not reviewed or $y < y_0$ and the case is reviewed. Thus

$$\begin{aligned} \Pr(y < y_0) &= \Pr\left(q_h < y_0 \text{ and } \left|g^{-1}(q_l) - q_h\right| > c\right) \\ &\quad + \Pr\left(g^{-1}(q_l) < y_0 \text{ and } \left|g^{-1}(q_l) - q_h\right| \le c\right) \end{aligned}$$

We may sum the two probabilities on the right-hand side because the events are disjoint. We consider each of these terms separately.

$$\begin{aligned} \Pr\left(q_h < y_0 \text{ and } \left|g^{-1}(q_l) - q_h\right| > c\right) &= \Pr\left(q_h < y_0 \text{ and } (q_l) \notin [q_h - c, q_h + c]\right) \\ &= \int_{-\infty}^{y_0} \left[1 - F^l(g(q_h + c)) + F^l(g(q_h - c))\right] f^h(q_h)\, dq_h \end{aligned}$$

And

$$\Pr\left(\begin{array}{c} g^{-1}(q_l) < y_0 \\ \text{and } \left|g^{-1}(q_l) - q_h\right| \le c \end{array}\right) = \Pr\left(q_l < g(y_0) \text{ and } (q_l) \in [q_h - c, q_h + c]\right)$$

$$= \int_{-\infty}^{y_0+c} \int_{g(q_h-c)}^{\min\{g(y_0),g(q_h+c)\}} f^h(q_h) f^l(q_l)\, dq_l dq_h$$

$$= \int_{-\infty}^{y_0-c} \int_{g(q_h-c)}^{g(q_h+c)} f^h(q_h) f^l(q_l)\, dq_l dq_h$$

$$+ \int_{y_0-c}^{y_0+c} \int_{g(q_h-c)}^{g(y_0),} f^h(q_h) f^l(q_l)\, dq_l dq_h$$

$$= \int_{-\infty}^{y_0-c} \left[F^l(g(q_h+c)) - F^l(g(q_h-c))\right] f^h(q_h)\, dq_h$$

$$+ \int_{y_0-c}^{y_0+c} \left[F^l(g(y_0)) - F^l(g(q_h-c))\right] f^h(q_h)\, dq_h$$

Hence

$$F^y(y_0) = \int_{-\infty}^{y_0} \left[1 - F^l(g(q_h+c)) + F^l(g(q_h-c))\right] f^h(q_h)\, dq_h$$

$$+ \int_{-\infty}^{y_0-c} \left[F^l(g(q_h+c)) - F^l(g(q_h-c))\right] f^h(q_h)\, dq_h$$

$$+ \int_{y_0-c}^{y_0+c} \left[F^l(g(y_0)) - F^l(g(q_h-c))\right] f^h(q_h)\, dq_h$$

By Leibnitz's Rule,

$$
\begin{aligned}
f^y(y) &= \Big[1 - F^l(g(y+c)) + F^l(g(y-c))\Big] f^h(y) \\
&\quad + \int_{y-c}^{y+c} f^l(g(y))\, g'(y)\, f^h(q_h)\, dq_h \\
&= \Big[1 - F^l(g(y+c)) + F^l(g(y-c))\Big] f^h(y) \\
&\quad + \Big[F^h(y+c) - F^h(y-c)\Big] f^l(g(y))\, g'(y)
\end{aligned}
$$

Lemma 1

There exists x_0 such that $d(x_0) = 0$.

Proof.

$$
\begin{aligned}
\int_{-\infty}^{\infty} d(x)dx &= \int_{-\infty}^{\infty} c\left[f^h(x+c) + f^h(x-c)\right] dx - \int_{-\infty}^{\infty}\int_{x-c}^{x+c} f^h(t)dtdx \\
&= 2c - 2c\int_{-\infty}^{\infty} f^h(t)dt \\
&= 0
\end{aligned}
$$

Since d is continuous, there must be some x_0 for which $d(x_0) = 0$. ■

Lemma 2

If f^h is convex on the interval $(x-c, x+c)$, then $d(x) > 0$. Similarly, if f^h is concave on the interval $(x-c, x+c)$, then $d(x) < 0$

Proof.

$$
\begin{aligned}
d(x) &= c\left[f^h(x+c) + f^h(x-c)\right] - \int_{x-c}^{x+c} f^h(t)dt \\
&= \int_{x-c}^{x+c} \left[\frac{x-l+c}{2c} f^h(x-c) + \frac{l-x+c}{2c} f^h(x+c) - f(t)\right] dt \\
&= \int_{x-c}^{x+c} \left[\frac{x-t+c}{2c} f^h(x-c) + \frac{t-x+c}{2c} f^h(x+c) - f^h\left(\frac{x-t+c}{2c}(x-c) + \frac{t-x+c}{2c}(x+c)\right)\right] dt \\
&> 0
\end{aligned}
$$

■

Lemma 3

$$\begin{aligned}\frac{d\lambda}{dc} &= -\frac{\left(c^2+c\lambda\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c^2-c\lambda\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}{\left(c\lambda+c^2+\sigma^2\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c\lambda-c^2-\sigma^2\right)\phi\left(\frac{\lambda-c}{\sigma}\right)} \text{ and} \\ \frac{d\lambda}{d\sigma} &= \frac{\left[\frac{c}{\sigma}(\lambda+c)^2+\sigma\lambda\right]\phi\left(\frac{\lambda+c}{\sigma}\right)+\left[\frac{c}{\sigma}(\lambda-c)^2-\sigma\lambda\right]\phi\left(\frac{\lambda-c}{\sigma}\right)}{\left(c\lambda+c^2+\sigma^2\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c\lambda-c^2-\sigma^2\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}.\end{aligned}$$

Proof.

$$\frac{c}{\sigma}\left[\phi\left(\frac{\lambda+c}{\sigma}\right)+\phi\left(\frac{\lambda-c}{\sigma}\right)\right]=\Phi\left(\frac{\lambda+c}{\sigma}\right)-\Phi\left(\frac{\lambda-c}{\sigma}\right)$$

Differentiating yields

$$\left(\begin{array}{c}\frac{c}{\sigma^2}\phi'\left(\frac{\lambda+c}{\sigma}\right)\left(\frac{d\lambda}{dc}+1\right) \\ +\frac{c}{\sigma^2}\phi'\left(\frac{\lambda-c}{\sigma}\right)\left(\frac{d\lambda}{dc}-1\right) \\ +\frac{1}{\sigma}\left[\phi\left(\frac{\lambda+c}{\sigma}\right)+\phi\left(\frac{\lambda-c}{\sigma}\right)\right]\end{array}\right)=\left(\begin{array}{c}\frac{1}{\sigma}\phi\left(\frac{\lambda+c}{\sigma}\right)\left(\frac{d\lambda}{dc}+1\right) \\ -\frac{1}{\sigma}\phi\left(\frac{\lambda-c}{\sigma}\right)\left(\frac{d\lambda}{dc}-1\right)\end{array}\right)$$

Rearranging terms and substituting for ϕ' yields

$$\frac{d\lambda}{dc}=-\frac{\left(c^2+c\lambda\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c^2-c\lambda\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}{\left(c\lambda+c^2+\sigma^2\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c\lambda-c^2-\sigma^2\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}$$

We similarly differentiate with repsect to σ:

$$\begin{aligned}&\frac{c}{\sigma}\phi'\left(\frac{\lambda+c}{\sigma}\right)\left[\frac{1}{\sigma}\frac{d\lambda}{d\sigma}-\frac{\lambda+c}{\sigma^2}\right]+ \\ &\frac{c}{\sigma}\phi'\left(\frac{\lambda-c}{\sigma}\right)\left[\frac{1}{\sigma}\frac{d\lambda}{d\sigma}-\frac{\lambda-c}{\sigma^2}\right]- \\ &\frac{c}{\sigma^2}\left[\phi\left(\frac{\lambda+c}{\sigma}\right)+\phi\left(\frac{\lambda-c}{\sigma}\right)\right] = \phi\left(\frac{\lambda+c}{\sigma}\right)\left[\frac{1}{\sigma}\frac{d\lambda}{d\sigma}-\frac{\lambda+c}{\sigma^2}\right] \\ &\qquad -\phi\left(\frac{\lambda-c}{\sigma}\right)\left[\frac{1}{\sigma}\frac{d\lambda}{d\sigma}-\frac{\lambda-c}{\sigma^2}\right]\end{aligned}$$

Rearranging terms yields

$$\frac{d\lambda}{d\sigma}=\frac{\left[\frac{c}{\sigma}(\lambda+c)^2+\sigma\lambda\right]\phi\left(\frac{\lambda+c}{\sigma}\right)+\left[\frac{c}{\sigma}(\lambda-c)^2-\sigma\lambda\right]\phi\left(\frac{\lambda-c}{\sigma}\right)}{\left(c\lambda+c^2+\sigma^2\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c\lambda-c^2-\sigma^2\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}$$

■

Proof of Theorem 5:

Part (a):

We construct Taylor series for Φ and ϕ around λ:

$$\begin{aligned}\frac{c}{\sigma}\left[\phi\left(\frac{\lambda+c}{\sigma}\right)+\phi\left(\frac{\lambda-c}{\sigma}\right)\right] &= \frac{2c}{\sigma}\phi\left(\frac{\lambda}{\sigma}\right)+\frac{c^3}{\sigma^3}\phi''\left(\frac{\lambda}{\sigma}\right)+O\left(c^4\right)\\ \Phi\left(\frac{\lambda+c}{\sigma}\right)-\Phi\left(\frac{\lambda-c}{\sigma}\right) &= \frac{2c}{\sigma}\phi\left(\frac{\lambda}{\sigma}\right)+\frac{c^3}{3\sigma^3}\phi''\left(\frac{\lambda}{\sigma}\right)+O\left(c^4\right)\end{aligned}$$

If λ satisfies Equation (*) for sufficiently small c, it follows that $\phi''\left(\frac{\lambda}{\sigma}\right)=0$, or equivalently, $\left[1-\left(\frac{\lambda}{\sigma}\right)^2\right]\phi\left(\frac{\lambda}{\sigma}\right)=0 \Longrightarrow \lambda=\sigma$.

Part (b):

$$\begin{aligned}\lim_{c\to 0}\frac{d\lambda}{dc} &= \lim_{c\to 0}-\frac{\left(c^2+c\lambda\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c^2-c\lambda\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}{\left(c\lambda+c^2+\sigma^2\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c\lambda-c^2-\sigma^2\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}\\ &= \lim_{c\to 0}-\frac{\lambda\left[\phi\left(\frac{\lambda+c}{\sigma}\right)-\phi\left(\frac{\lambda-c}{\sigma}\right)\right]}{\lambda\left[\phi\left(\frac{\lambda+c}{\sigma}\right)+\phi\left(\frac{\lambda-c}{\sigma}\right)\right]+\frac{\sigma^2}{c}\left[\phi\left(\frac{\lambda+c}{\sigma}\right)-\phi\left(\frac{\lambda-c}{\sigma}\right)\right]}\\ &= \frac{0}{2\lambda\phi\left(\sigma\right)+\sigma^2\phi'\left(\sigma\right)}=0\end{aligned}$$

Part (c):

$$\begin{aligned}\lim_{c\to\infty}\frac{d\lambda}{dc} &= \lim_{c\to\infty}-\frac{\left(c^2+c\lambda\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c^2-c\lambda\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}{\left(c\lambda+c^2+\sigma^2\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c\lambda-c^2-\sigma^2\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}\\ &= \lim_{c\to\infty}-\frac{\left(c+\lambda\right)\frac{\phi\left(\frac{\lambda+c}{\sigma}\right)}{\phi\left(\frac{\lambda-c}{\sigma}\right)}+\left(c-\lambda\right)}{\left(\lambda+c+\frac{\sigma^2}{c}\right)\frac{\phi\left(\frac{\lambda+c}{\sigma}\right)}{\phi\left(\frac{\lambda-c}{\sigma}\right)}+\left(\lambda-c-\frac{\sigma^2}{c}\right)}\\ &= \lim_{c\to\infty}-\frac{\left(c+\lambda\right)e^{\frac{-2c\lambda}{\sigma^2}}+\left(c-\lambda\right)}{\left(\lambda+c\right)e^{\frac{-2c\lambda}{\sigma^2}}+\left(\lambda-c\right)}\end{aligned}$$

Note that we cannot have $\lambda\to 0$ as $c\to\infty$; if this were true, then $\frac{2c}{\sigma}\phi\left(\frac{c}{\sigma}\right)\to 1$ as $c\to\infty$, which is impossible. Hence $e^{\frac{-2c\lambda}{\sigma^2}}\to 0$ as $c\to\infty$, so

$$\lim_{c\to\infty}\frac{d\lambda}{dc}=\frac{\lambda-c}{\lambda-c}=1.$$

Part (d):

From Lemma 3,

$$\begin{aligned}
\frac{d\lambda}{dc} &= -\frac{\left(c^2+c\lambda\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c^2-c\lambda\right)\phi\left(\frac{\lambda-c}{\sigma}\right)}{\left(c\lambda+c^2+\sigma^2\right)\phi\left(\frac{\lambda+c}{\sigma}\right)+\left(c\lambda-c^2-\sigma^2\right)\phi\left(\frac{\lambda-c}{\sigma}\right)} \\
&= \frac{\frac{c^2}{\sigma^2}-\left(\frac{c\lambda}{\sigma^2}\right)\tanh\frac{c\lambda}{\sigma^2}}{\left(\frac{c^2}{\sigma^2}+1\right)\tanh\frac{c\lambda}{\sigma^2}-\left(\frac{c\lambda}{\sigma^2}\right)}
\end{aligned}$$

First, we consider the case where $\sigma = 1$, so that

$$\frac{d\lambda}{dc} = \frac{c^2-(c\lambda)\tanh c\lambda}{(c^2+1)\tanh c\lambda-(c\lambda)} \tag{1}$$

From part (a), we know that $\lambda \to 1$ as $c \to 0$. Substituting the power series of tanh into (1) yields

$$\frac{d\lambda}{dc} = \frac{c^2(1-\lambda^2)+\frac{1}{3}c^4\lambda^4+O(c^6)}{c^3(\lambda-\frac{1}{3}\lambda^3)+O(c^5)}$$

so that as $c \to 0$, $\frac{d\lambda}{dc} \to 0$ and $\frac{d^2\lambda}{dc^2} \to \frac{1}{2}$. Thus λ is increasing in c when c is sufficiently close to zero.

Now consider the curve S defined by the equation

$$c^2-(c\lambda)\tanh c\lambda = 0 \tag{2}$$

Implicitly differentiating (2) yields

$$\frac{d\lambda}{dc} = \frac{1-c^2\operatorname{sech}^2 c\lambda}{\tanh c\lambda + c\lambda\operatorname{sech}^2 c\lambda} > 0 \tag{3}$$

so S has strictly positive slope. Note that S contains the point $(c,\lambda) = (0,1)$, since $\lambda \to 1$ as $c \to 0$ in equation (2). Also, as can be seen in equation (3), S has infinite slope at $(0,1)$. Thus, for sufficiently small values of c, the solution to (1) must lie between the x-axis and curve S. If the solution to (1) intersected S at another point, it must approach S from below, and hence have positive slope at that point. However, the solution to (1) must have zero slope at any point at which it intersects S. Therefore the solution to (1) will always be between the x-axis and curve S.

In the region between the x-axis and curve S, the numerator $\left[c^2 - (c\lambda)\tanh c\lambda\right]$ in (1) will be strictly positive, and the denominator will be postive for all $c, \lambda > 0$. Hence $\frac{d\lambda}{dc}$ is strictly positive, so λ is strictly increasing in c.

The general case, when $\sigma \neq 1$, now follows easily. Let $\widetilde{c} = \frac{c}{\sigma}$, $\widetilde{\lambda} = \frac{\lambda}{\sigma}$, and $\widehat{\sigma} = 1$. By the above reasoning, $\frac{d\widehat{\lambda}}{d\widehat{c}}$ is strictly increasing, hence $\frac{d\lambda}{dc}$ is strictly increasing.

Lemma 4

As $c \to 0$, $g(x) \to x + \frac{3}{2}\frac{\sigma^2 x}{\sigma^2 - x^2}$, and the density function of x approaches $\frac{1}{\sqrt{2\pi}\sigma}\left(1 + \frac{3}{2}\frac{\sigma^2(\sigma^2+x^2)}{(\sigma^2-x^2)^2}\right)\exp\left(-\frac{1}{2}x^2\left(1+\frac{3}{2}\frac{\sigma^2}{\sigma^2-x^2}\right)^2\right)$.

Proof. We use the Taylor Series expansions for $F^h(x) = \Phi\left(\frac{x}{\sigma}\right)$ and $f^h(x) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)$, substituting them into the formula for $g(x)$. This yields

$$g(x) = x + \frac{\sigma}{2}\frac{-2c^2\frac{x}{\sigma} + O(c^2)}{\frac{2}{3}c^2\left(\frac{x^2}{\sigma^2} - 1\right) + O(c^4)} \to x + \frac{3}{2}\frac{\sigma^2 x}{\sigma^2 - x^2}$$

Now $q_l = h(x)$, so as $c \to 0$,

$$\begin{aligned} f^x(x) &= f^l(g(x))g'(x) \\ &\to \phi\left(x + \frac{3}{2}\frac{\sigma^2 x}{\sigma^2 - x^2}\right)\left(1 + \frac{3}{2}\frac{\sigma^2\left(\sigma^2 + x^2\right)}{\left(\sigma^2 - x^2\right)^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\left(1 + \frac{3}{2}\frac{\sigma^2\left(\sigma^2 + x^2\right)}{\left(\sigma^2 - x^2\right)^2}\right)\exp\left\{-\frac{1}{2}x^2\left(1 + \frac{3}{2}\frac{\sigma^2}{\sigma^2 - x^2}\right)^2\right\} \end{aligned}$$

■

Lemma 5

As $c \to 0$, $f^x(x)$ will have a single peak at $x = 0$.

Proof. First, assume $\sigma^2 = 1$, so that

$$f^x(x) = \frac{1}{\sqrt{2\pi}\sigma}\left(1 + \frac{3}{2}\frac{\left(1 + x^2\right)}{\left(1 - x^2\right)^2}\right)\exp\left\{-\frac{1}{2}x^2\left(1 + \frac{3}{2}\frac{1}{1 - x^2}\right)^2\right\}$$

and let $s = \frac{\sigma^2}{\sigma^2 - x^2}$. Since x is defined on $[-1, 1]$, s will be defined on $[1, \infty]$. Substituting s into the above equation,

$$\log f^x(x) = -\log\sqrt{2\pi}\sigma + \log\left(1 + \frac{3}{2}s\left(2s - 1\right)\right) - \frac{1}{2}\left(1 - \frac{1}{s}\right)\left(1 + \frac{3}{2}s\right)^2$$

Taking derivatives yields

$$\frac{\partial}{\partial s}\log f^x(x) = \frac{12s-3}{6s^2-3s+2} - \frac{1}{2}\left[\frac{1}{s^2} + \left(3-\frac{9}{4}\right) + \frac{9}{2}s\right]$$

and

$$\begin{aligned}\frac{\partial^2}{\partial s^2}\log f^x(x) &= \frac{6-\frac{9}{4}+9\left(s-2s^2\right)}{\left(6s^2-3s+2\right)^2} + \left[\frac{1}{s^3}-\frac{9}{2}\right] \\ &\leq \frac{6-\frac{9}{4}v^2+9\left(-1\right)}{\left(6s^2-3s+2\right)^2} + \left[1-\frac{9}{2}\right], \text{ since } s \geq 1 \\ &\leq \frac{6-\frac{45}{4}}{\left(6s^2-3s+2\right)^2} + \left[1-\frac{9}{2}\right] \\ &< 0\end{aligned}$$

Thus $\frac{\partial}{\partial s}\log f^x(x)$ will be negative everywhere if $\frac{\partial}{\partial s}\log f^x(x) < 0$ at $s = 1$. Now

$$\left.\frac{\partial}{\partial s}\log f^x(x)\right|_{s=1} = \frac{9}{5} - \frac{25}{8} < 0$$

which provides the desired result. Also, $\frac{\partial s}{\partial x} > 0 \Leftrightarrow x > 0$, so $\frac{\partial}{\partial x}\log f^x(x) < 0 \Leftrightarrow x > 0$. Thus $f^x(x)$ is decreasing for $x > 0$ and increasing for $x < 0$. Therefore f^x has a single peak at $x = 0$ when $\sigma^2 = 1$.

Now for $\sigma^2 \neq 1$, we normalize the other parameters. Let $\widetilde{c} = \frac{c}{\sigma}$, $\widetilde{x} = \frac{x}{\sigma}$, and $\widetilde{\sigma} = 1$. Using the above reasoning, the density function of $\widetilde{x}$ must have a single peak at $\widetilde{x} = 0$. Since $x = \sigma\widetilde{x}$, it follows that the density of x must also have a single peak at $x = 0$. ■

The above results describe the shape of f^x, the density function of the lower court's ruling, as $c \rightarrow 0$. For sufficiently large values of σ, f^x is single-peaked and symmetric around 0, and bounded at $\pm\sigma$. It follows, and we state without proof, that the variance of x must be strictly less than σ. This means that as $c \rightarrow 0$, the lower court's rulings will have lower variance than the higher court's, irrespective of the distribution of both courts' preferences. However, as $c \rightarrow 0$, the likelihood that the lower court's ruling will stand approaches 0. For very small c, the second effect dominates, so that a slight increase in c will reduce the variance of the final ruling. This provides intuition for the following result.

Proof of Theorem 6:

When $c = 0$, every case will be appealed, so the final ruling will be the appeals court's preference point. Thus the distribution of the final ruling will be the same as the distribution of the appeals court's preferences, in this case, a normal distribution with mean 0 and variance σ^2. Now consider $\varepsilon > 0$. Let $v_0 = \sigma^2$ be the variance of the final ruling when $c = 0$, and v_ε be the variance when $c = \varepsilon$. Then the final ruling will be the same in both these cases unless $|x - q_h| \leq \varepsilon$. Thus

$$\begin{aligned} v_0 - v_\varepsilon &= \int_{-\infty}^{\infty} \int_{q_h-\varepsilon}^{q_h+\varepsilon} \left(q_h^2 - x^2\right) f^x(x) \frac{1}{\sigma} \phi\left(\frac{q_h}{\sigma}\right) dx dq_h \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma} \phi\left(\frac{q_h}{\sigma}\right) \underbrace{\int_{q_h-\varepsilon}^{q_h+\varepsilon} \left(q_h^2 - x^2\right) f^x(x) dx dq_h}_{\text{positive}} \end{aligned}$$

Note that the second integral is positive, because

$$\begin{aligned} \int_{q_h-\varepsilon}^{q_h+\varepsilon} \left(q_h^2 - x^2\right) f^x(x) dx &= \int_0^{\varepsilon} \left\{ [q_h^2 - (q_h + t)^2] f^x(q_h + t) + [q_h^2 - (q_h - t)^2] f^x(q_h - t) \right\} dt \\ &= \int_0^{\varepsilon} \left\{ [-2q_h t - t^2] f^x(q_h + t) dt + [2q_h t - t^2] f^x(q_h - t) \right\} dt \\ &= \underbrace{\int_0^{\varepsilon} 2q_h t [f^x(q_h - t) - f^x(q_h + t)] dt}_{\text{postive, } O(\varepsilon^2)} + \underbrace{\int_0^{\varepsilon} -t^2 [f^x(q_h - t) + f^x(q_h + t)] dt}_{\text{negative, } O(\varepsilon^3)} \end{aligned}$$

so it follows that the entire double integral is positive. Therefore $v_0 > v_\varepsilon$ for small ε.

Proof of Theorem 7:

Part (a): Let $\widetilde{\lambda} = \frac{\lambda}{\sigma}$ and $\widetilde{c} = \frac{c}{\sigma}$. Then $\widetilde{c}\left[\phi\left(\widetilde{\lambda} + \widetilde{c}\right) + \phi\left(\widetilde{\lambda} - \widetilde{c}\right)\right] = \Phi\left(\widetilde{\lambda} + \widetilde{c}\right) - \Phi\left(\widetilde{\lambda} - \widetilde{c}\right)$. Now as $\sigma \to 0$, $\widetilde{c} \to \infty$. From Theorem 5, $\frac{\widetilde{\lambda}}{\widetilde{c}} \to 1$ as $\widetilde{c} \to \infty$, hence $\frac{\lambda}{c} \to 1$. Thus $\lim_{\sigma \to 0} \lambda(c, \sigma) = c$.

Part (b): As above, let $\widetilde{\lambda} = \frac{\lambda}{\sigma}$ and $\widetilde{c} = \frac{c}{\sigma}$, so that $\widetilde{c}\left[\phi\left(\widetilde{\lambda} + \widetilde{c}\right) + \phi\left(\widetilde{\lambda} - \widetilde{c}\right)\right] = \Phi\left(\widetilde{\lambda} + \widetilde{c}\right) - \Phi\left(\widetilde{\lambda} - \widetilde{c}\right)$. In this case, we have $\widetilde{c} \to 0$ as $\sigma \to \infty$. Thus $\widetilde{\lambda} \to 1$, so $\frac{\lambda}{\sigma} \to 1$ as $\sigma \to \infty$. Hence $\lim_{\sigma \to \infty} \frac{d\lambda}{d\sigma} = 1$.

Part (c): First, we must show that $\lim_{\sigma \to 0} \frac{\lambda - c}{\sigma} = -\infty$. Since $\frac{c}{\sigma}\left[\phi\left(\frac{\lambda + c}{\sigma}\right) + \phi\left(\frac{\lambda - c}{\sigma}\right)\right] = \Phi\left(\frac{\lambda + c}{\sigma}\right) - \Phi\left(\frac{\lambda - c}{\sigma}\right)$, it follows that $\phi\left(\frac{\lambda - c}{\sigma}\right) < \frac{\sigma}{c}$. Thus $-\frac{1}{2}\left(\frac{\lambda - c}{\sigma}\right)^2 < \log \frac{\sigma\sqrt{2\pi}}{c} \implies \left|\frac{\lambda - c}{\sigma}\right| > \sqrt{\log \frac{c^2}{2\pi\sigma^2}}$. Note that $\frac{\lambda - c}{\sigma} > \sqrt{\log \frac{c^2}{2\pi\sigma^2}}$ is impossible for large enough σ, since $\phi()$ will then be convex on $\left(\frac{\lambda-c}{\sigma}, \frac{\lambda+c}{\sigma}\right)$. Hence, $\frac{\lambda-c}{\sigma} < -\sqrt{\log\frac{c^2}{2\pi\sigma^2}}$. Thus, as $\sigma \to 0$, $\frac{\lambda-c}{\sigma} \to -\infty$.

Now,

$$
\begin{aligned}
\lim_{\sigma\to 0}\frac{d\lambda}{d\sigma} &= \lim_{\sigma\to 0}\frac{\left[\frac{c}{\sigma}(\lambda+c)^2+\sigma\lambda\right]\phi\left(\frac{\lambda+c}{\sigma}\right)+\left[\frac{c}{\sigma}(\lambda-c)^2-\sigma\lambda\right]\phi\left(\frac{\lambda-c}{\sigma}\right)}{(c\lambda+c^2+\sigma^2)\,\phi\left(\frac{\lambda+c}{\sigma}\right)+(c\lambda-c^2-\sigma^2)\,\phi\left(\frac{\lambda-c}{\sigma}\right)} \\
&= \lim_{\sigma\to 0}\frac{\left[\frac{c}{\sigma}(\lambda+c)^2+\sigma\lambda\right]e^{\frac{-2c\lambda}{\sigma^2}}+\left[\frac{c}{\sigma}(\lambda-c)^2-\sigma\lambda\right]}{(c\lambda+c^2+\sigma^2)\,e^{\frac{-2c\lambda}{\sigma^2}}+(c\lambda-c^2-\sigma^2)} \\
&= \lim_{\sigma\to 0}\frac{\left[c\left(\frac{\lambda-c}{\sigma}\right)^2-\lambda\right]}{\left(c\left(\frac{\lambda-c}{\sigma}\right)-\sigma\right)} \\
&= \lim_{\sigma\to 0}\frac{\lambda-c}{\sigma} = -\infty.
\end{aligned}
$$

Proof of Theorem 8:

The bounds of rulings will be invariant to d because they are determined by where the denominator of $g(x)$ crosses the x-axis. Since d does not appear in the denominator of $g(x)$, it will not affect the bounds of rulings.

To show that the variance of x is decreasing in d, note that

$$
g(x) = x + \frac{1}{2}\frac{\left(\frac{c^2+d}{\sigma}\right)\left[\phi\left(\frac{x+c}{\sigma}\right)-\phi\left(\frac{x-c}{\sigma}\right)\right]}{\frac{c}{\sigma}\left[\phi\left(\frac{x+c}{\sigma}\right)+\phi\left(\frac{x-c}{\sigma}\right)\right]-\left[\Phi\left(\frac{x+c}{\sigma}\right)-\Phi\left(\frac{x-c}{\sigma}\right)\right]}.
$$

By theorem 2, the fractional part above will be positive when $x > 0$ and negative when $x < 0$. Thus $\left|\frac{\partial g}{\partial x}\right|$ is strictly increasing in d, and therefore $\left|\frac{\partial g^{-1}}{\partial x}\right|$ is strictly decreasing in d. Since $x = g^{-1}(q_l)$, this means that a larger d results in lower variance of x.

References

ABA Commission on Immigration Policy (2003). *Seeking Meaningful Review: Findings and Recommendations in Response to Dorsey and Whitney Study of Board of Immigration Appeals Procedural Reforms*. ABA Commission on Immigration Policy.

Banks, C. P. (1997). The politics of en banc review in the "mini-supreme court". *Journal of Law and Politics 13*, 377–414.

Cameron, C. M. and L. A. Kornhauser (2005). Decision rules in a judicial hierarchy. *Journal of Institutional and Theoretical Economics 161*, 264–292.

Cameron, C. M., J. A. Segal, and D. Songer (2000). Strategic auditing in a political hierarchy: An informational model of the supreme court's certiorari decisions. *American Political Science Review 94*(1), 101–116.

Daughety, A. F. and J. Reinganum (2000). Appealing judgments. *RAND Journal of Economics 31*(3), 502–525.

Epstein, L. and J. Knight (1998). *The Choices Justices Make*. Washington, D.C.: Congressional Quarterly, Inc.

Epstein, L., J. A. Segal, H. J. Spaeth, and T. G. Walker (1996). *The Supreme Court Compendium*. Washington, D.C.: Congressional Quarterly, Inc.

Ferejohn, J. and C. Shipan (1990). Congressional influence on bureaucracy. *Journal of Law, Economics, and Organization 6*, 1–20.

Ginsburg, R. B. and P. W. Huber (1987). The intercircuit committee. *Harvard Law Review 100*, 1417–1435.

Haire, S. B., S. A. Lindquist, and D. R. Songer (2003). Appellate court supervision in the federal judiciary: A hierarchical perspective. *Law and Society Review 37*(1), 143–167.

Hellman, A. D. (1996). The shrunken docket of the rehnquist court. *Supreme Court Review 1996*, 403–438.

Klein, D. E. and R. J. Hume (2003). Fear of reversal as an explanation of lower court compliance. *Law and Society Review 37*(3), 579–606.

McNollgast (1995). Politics and the courts: A positive theory of judicial doctrine and the rule of law. *Southern California Law Review 68*, 1631–1683.

Merritt, D. J. and J. J. Brudney (2001). Stalking secret law: What predicts publication in the united states courts of appeals. *Vanderbilt Law Review 54*, 71–121.

Mialon, H. M., P. H. Rubin, and J. L. Schrag (2004). Judicial hierarchies and the rule-individual tradeoff. Law and Economics Research Paper 05-5, Emory.

O'Scannlain, D. F. (1999). Should the ninth circuit be saved? *Journal of Law and Politics 15*, 415–443.

Procter Hug, J. (2000). A preferable approach for the ninth circuit. *California Law Review 88*, 1657–1672.

Segal, J. A. and H. J. Spaeth (2002). *The Supreme Court and the Attitudinal Model Revisited*. United Kingdom: Cambridge University Press.

Shavell, S. (1995). The appeals process as a means of error correction. *Journal of Legal Studies 24*(2), 379–426.

Spiller, P. T. and M. L. Spitzer (1995). Where is the sin in sincere? sophisticated manipulation of sincere judicial voters (with applications to other voting environments). *Journal of Law, Economics, and Organization 11*(1), 32–63.

Spitzer, M. and E. Talley (2000). Judicial auditing. *Journal of Legal Studies 29*(2), 649–683.

Sunstein, C. R., D. Schkade, and L. M. Ellman (2004). Ideological voting on federal courts of appeals: A preliminary investigation. *Virginia Law Review 90*(1), 301–353.

William N. Eskridge, J. and J. Ferejohn (1992). Making the deal stick: Enforcing the original constitutional structure of lawmaking in the modern regulatory state. *Journal of Law, Economics, and Organization 8*(1), 165–189.