Appellate Lawmaking in a Judicial Hierarchy*

Adam B. Badawi † and Scott Baker ‡

† School of Law, Washington University in St. Louis
‡ School of Law, Washington University in St. Louis

August 8, 2014

Abstract

We examine an infinite horizon model of appellate court lawmaking. The model focuses on the impact of lawmaking on the behavior of trial courts, incorporating costly review of trial court dispositions and studying the properties of vertical stare decisis over time. A successful audit provides an opportunity to create precedent. Precedent changes future trial court dispositions by providing new cases from which trial courts can draw analogies. This, in turn, alters the appellate court's scrutiny of these dispositions – its audit strategy – going forward. We use the model to explain many features of actual judicial practice by courts in a hierarchy. Throughout, we relate the findings to existing evidence and derive testable predictions.

*For valuable feedback, we thank the editor, Richard Holden, Albert Choi, Lewis Kornhauser, William Hubbard, David Law, Claudio Mezzetti, Judge Richard Posner and participants at the workshop at Washington University School of Law, University of Virginia Law School, the 2014 Annual Meeting of the American Law and Economics Association and the 2014 meetings of the International Society for New Institutional Economics.

> Given that (1) Defendant [Officer] Leveille did not assert authority over Plaintiff beyond requesting that she leave, (2) Plaintiff did not in any way submit to his request, and (3) Plaintiff indisputably "retrieved Mr. Byers's belongings and walked away" immediately after the incident, this case is on all fours with McCoy v. Harrison, 341 F.3d 600, 605–06 (7th Cir. 2003).[1]

1 Introduction

The quote above is from a district court judge. It reflects analysis that is commonplace for lawyers and trial court judges, but rarely examined formally. The district court judge looked at the facts the case presented, compared those facts to the facts in an existing appellate court decision, decided the appellate court precedent was very similar to the case at hand ("on all fours") and, as a result, ruled in the officer's favor. The existence of a closely analogous precedent made it easier for the district court judge to resolve the case.

To capture the dynamics of this kind of reasoning, we present a game theoretic model of a judicial hierarchy where distance between a new case and precedent matters. Specifically, the lower court's cost of legal reasoning depends on the distance between the case at hand and the existing appellate court precedent. By issuing new precedent, the appellate court transforms "hard" cases into "easy" ones for trial courts. This transformation alters how trial courts decide cases. It makes one resolution – say, liable – easier to justify by reference to appellate court precedent than the alternative resolution – not liable. Locating the new precedent appropriately allows the appellate court to change future trial court decisions in a direction it favors. As a result, the the appellate court can audit trial courts less intensely, saving its own resources.

By including reasoning costs in the trial court's utility function, our model can explain many common judicial practices that other models cannot. Take a model where the appellate court reverses the trial court any time it perceives an error and consider the frequent practice of an appellate court writing a lengthy opinion that affirms the result of the lower court. An appellate court motivated by error correction would not expend the effort necessary to author this opinion. Once it perceives no error, it can simply affirm with a one-line order.

[1] District Court Judge Dow, Lewis v. City of Chicago (April 28, 2014) (summary judgment decision in favor of Officer Leveille and the City of Chicago).

Courts do not, however, proceed in this way. Appellate courts will often take the time to explain, in detail, the facts of the lower court case and why those facts lead to the ultimate result. Our model shows the benefits of doing so. By producing this new precedent, the appellate court allows trial courts that care about reasoning costs to resolve future cases more easily in a way that the appellate court prefers.

Our account also resonates with the actual content of trial court and appellate court opinions. Trial courts often speak of whether a case is "on point" or is on "all fours" with existing precedent. Where those cases do not exist, the trial court will struggle to reach a result. And it will do so with the knowledge that it is much more likely to be reversed in that sort of case than in the one where it can apply closely analogous precedent. That type of analysis is what many trial courts do in our model.

Next consider the persistent, but puzzling, judicial practice of including dicta in opinions. Appellate courts are traditionally reluctant to include any information about how they would resolve cases that are not before them. But by including dicta, appellate courts often do just that. We explain this strategy as a way to lower the decision costs of lower courts. When an appellate court has an actual case in front of it, that court may be able to say something useful about closely related hypothetical cases. But the usefulness of those statements to lower courts decreases as their distance from the actual case increases. Our model is consistent with the familiar belief among lawyers and legal scholars that statements that are unconnected from a concrete controversy do little to help the resolution of future cases. But a hypothetical that slightly modifies the facts of an actual case can be illuminating for future trial courts that consider similar cases. We thus characterize dicta as a maximization problem; the appellate court wants to set dicta at a level that optimizes the lower court's use of those statements. We show that the use of dicta makes the long-lived appellate court strictly better off.

In addition to explaining common practices, the model makes three predictions. First, and quite intuitively, we predict that appellate courts will issue opinions at different rates over time. These rates will depend on the size of the existing precedent stock. The key is whether enough "law" exists from which the trial court can look for assistance in its legal reasoning. If so, the appellate court doesn't have to do much auditing to ensure compliance with its

preferences.[2]

Second, we predict that the appellate courts will review trial court dispositions that it disagrees with more intensely than those that it favors. One reason for this result is obvious: the appellate court suffers a loss from a disposition with which it might disagree; a grant of a habeas petition perhaps. But our model suggests a less obvious reason as well. Dispositions that the appellate court might disagree with provide a better opportunity to make new law than dispositions that the appellate court views favorably. Imagine, as an example, that the appellate court must determine the amount of precaution that is necessary to escape liability for a tort. Further suppose that the appellate court wants to limit a prior case finding the defendant not liable to its facts. The best case to do so is an analogous case located right next to that precedent. If the appellate court issues an opinion in such a case finding the defendant liable, that liable precedent will impact a large number of subsequent cases in the lower courts. Given the prior not liable precedent on the books (the one whose reach the appellate court wishes to limit), the trial court is apt to find the defendant not liable in any closely analogous case. A finding of no liability thus signals to the appellate court that the case is closely analogous and, as a result, a good candidate for reversing and limiting the prior holding. If instead the trial court has found the defendant liable, the appellate court can infer that the case is unlikely to lie close to the not liable precedent and thus be a less effective vehicle for making law.

Finally, we dynamically model how appellate courts are likely to scrutinize lower courts that are solely motivated by legalist values and those that act on policy preferences alone. Legalist values capture those that are internal to the legal system such as the importance of following precedent and the method of reasoning by analogy. Realist concerns, on the other hand, involve preferences for external policy outcomes. Unremarkably, we predict that appellate scrutiny of these two categories of trial courts differs.

When the law converges, the appellate court continues will continue to scrutinize decisions by realist trial courts, but will defer to dispositions by legalist

[2] In this way, our model combines the "enforcement" function and "doctrinal creation" functions of appellate courts. It responds to a call from Cameron et al. (2000) in their paper on the grant of certiorari. They state "the incremental and fact-soaked creation of new rules is one of the most interesting and distinctive elements of judicial politics." Id. at 102. Their model does not consider the issue of how the Supreme Court creates doctrine, while anticipating the need to review future decisions by lower courts. They note that "law creation is an important part of certiorari, but modeling it involves even more complex theoretical issues than the enforcement of doctrine." Id. at 103.

trial courts. The appellate court continues to suffer losses: it doesn't get its preferred outcome in every case. That happens as a byproduct of delegating initial decisionmaking to an agent (the realist trial court) with conflicting preferences. Importantly, the appellate court losses from this delegation are smaller when it makes law than when it doesn't. Lawmaking, in other words, helps to reduce the agency costs associated with the delegation of initial decisionmaking to lower courts.

The next subsection reviews the related literature. Part 2 is a short numerical example. It serves to illustrate many of the themes of the formal model. Section 3 lays out a one-period model. It benchmarks the dynamic model that comes in Section 4. That section derives the results regarding how the appellate court will make law and review trial court dispositions over time. Section 5 studies dicta. Section 6 provides a discussion of the assumptions underlying the modeling exercise. A short conclusion follows. The appendix contains all proofs. Throughout, we identify testable predictions of the model and suggest consistencies between our model and the available evidence.

1.1 Related Literature

Our project relates to two existing literatures. The first examines an "enforcement" game between upper and lower courts (see, for example, McNollGast (1995); Cameron et al. (2000); Spitzer and Talley (2000); Shavell (2006); Lax (2012); Beim et al. (2014)). The principal is the upper court and the lower courts are agents. The upper and lower court disagree about the proper way to resolve cases. On occasion, the upper court has more information about the content of the law than the lower court (Daughety and Reinganum (2000), for example). Subject to limited resources, the upper court reverses lower court decisions it disagrees with or that are erroneous (Songer et al. (1995); McNollGast (1995); Spitzer and Talley (2000); Daughety and Reinganum (2000)). The trial court anticipates the threat of reversal in making its decisions.[3]

Some of this literature adds wrinkles to the enforcement game, producing more insights. In McNollgast, for example, precedent demarcates the kinds of

[3]The rationale law professors offer for why trial courts follow precedent does not always depend on reversal-avoidance. Some accounts assume that lower court judges intrinsically care about following precedent and have preferences over policy outcomes (Cross (2005); Kim (2007)). The issue, then, is which concern dominates in a particular case. Unsurprisingly, assuming lower courts care about following precedent generates a predication that they will tend to follow precedent. But why do lower court judges care about appellate court precedent? What function does it serve?

lower court decisions that will be potentially subject to appellate scrutiny. In Lax (2012), the appellate court decides whether to formulate precedent as rule or a standard, where a standard allows more room for the trial court to engage in non-compliance. Gennaioli and Schliefer (2008) treat the determination of fact as a choice variable of the trial courts. In that model, the trial court can choose to report facts in such a way as to avoid appellate review.[4]

This literature considers a single interaction between the upper and lower courts. Given this assumption, this literature cannot account for changes in precedent over time. It does not explain why appellate courts build precedent incrementally or why changing the precedent today helps the appellate court ensure that trial courts make decisions more consistent with the appellate court's preferences tomorrow. Our model tackles these issues directly.

A second related literature examines the role of precedent as between judges in the same tier – the issue of horizontal stare decisis. Why do Supreme Court judges follow their own precedent? Why do circuit judges follow the precedent created by other circuit judges? This literature provides some answers. Appellate judges in Gennaioli and Schliefer (2007) search for distinctions between their case and the prior precedent. They follow precedent when it is too costly to find a distinguishing feature in the case before them. The quest to distinguish improves the overall performance of the law. Baker & Mezzetti (2012) show that, given limited resources, a long-lived, imperfectly informed appellate court will always follow precedent. In that model, precedent reveals information about the way case closes to the precedent should be decided. Rather than spend resources taking a fresh look, the appellate court summarily decides those cases by reference to precedent. In Stephenson and Bueno De Mesquita (2002), the appellate court judge often follows precedent with which he disagrees. He does so because adding new precedent to the stock of existing precedent improves communication between himself and the trial judge. In O-Hara (1993) and Rasmusen (1994), respect for precedent arises as an equilibrium of a repeated game among judges. Appellate judges follow precedent created by other appellate judges, so that future judges follow the precedent they create. These papers take the flow of cases as given and consider how appellate judges react to their own appellate court precedent.

In contrast, our paper asks how precedent building by the appellate court

[4]We do not incorporate the important issue of fact manipulation in this model. We speculate in the penultimate section of the paper about how such manipulation might play out in our framework.

affects the behavior of lower courts: the issue of vertical stare decisis. Why – and when – do lower courts follow appellate court precedent and how do changes in that precedent alter the appellate court's need for auditing?

2 Numerical Example

This section works through a numerical example that we base on the model. The context for this example is a socially beneficial activity that can be harmful to others if a party does not take sufficient precaution – a tort. When a lawsuit results, courts must decide whether the amount of precaution taken by the defendant should result in liability or no liability. Imagine that we line up the cases from the safest (i.e. the defendant took extreme precaution) to the most dangerous (i.e. the defendant took no precaution).[5] So, case 0 represents the safest possible activity and Case 1 represents the most dangerous activity. Case .5 is halfway in between. In the example there are seven possible cases and each has 1/7 chance of occurring. Figure 1 shows each of the cases.

The appellate court inherits some existing precedent, either from the Supreme Court or from a prior decision of the appellate court. This inherited precedent says that if a defendant takes a level of precaution at or below .1, that defendant should not be found liable. It also says that the activities above .95 should result in liability. There is no guidance on the cases in between .1 and .95. It is in this range that the appellate court will fulfill its traditional role of developing the law (Frisch, 2003).

Following much of the literature on judicial decision making, we assume that appellate judges have an ideal point in the case space (Kornhauser (1992a); Kornhauser (1992b); Lax (2007)). In this example, we imagine an appellate court that wishes to restrict the reach of the no liability precedent only to cases .1 and below. In other words, it wants defendants to be liable for all activities that are more dangerous than .1. We make the standard assumption that the appellate court suffers a loss if a case is resolved against its preferences (McNollgast, 1995). The question is how the appellate court will minimizes its losses, given that it has scarce resources and there is a cost to reviewing trial court dispositions.

[5] The ordering is from activities which set out the weakest case for liability to those with the strongest case for liability. It can be thought of in many ways other than as the degree of recklessness of the defendant. One could, for example, line up police searches from least intrusive to most intrusive or trademark infringement cases from ones where the plaintiff's mark was the weakest to ones where the plaintiff's mark was the strongest.

To answer that question we need to understand how the trial court will behave. For the purpose of the example, we look only at a "legalist" trial court.[6] This type of court has no preferences over outcomes. Instead, this trial court cares about effort it must devoted to analogical or legal reasoning. This approach to modeling trial court behavior follows the traditional principles of legal reasoning; it is generally considered more persuasive to make a close analogy than a stretched one (Sunstein, 1993). As lawyers sometimes put it, it is better to have a case that is "on point." We capture this effect through the effort that a trial court must make to make an analogy. It takes less effort to associate a new set of facts with a close case than it does to associate the new case with a distant case. Applying this approach, the legalist trial court resolves cases 0,.1, .35 and .5 in favor of the defendant (no liability). For these cases, the not liable precedent located at .1 is closer than the liable precedent located at .95. The legalist trial court resolves cases .7, .95 and 1 in favor of the plaintiff (liability). For these cases, the liable precedent located at .95 is closer than the not liable precedent located at .1.

This legalist approach is consistent with descriptions that judges give of their behavior. As Judge Posner explains: "in all cases of reasoning by analogy, sound analysis requires attending to the policy considerations that align the case at hand with one or another line of precedents." (Posner, 2006, p.766). Our model of the legalist trial court attempts to capture this process. Often, the trial court will be confronted with a case with precedent on both sides, but no precedent directly on point. To reach one result or the other requires that the trial court reason by analogy from the precedent. All the model says is that the closer the case is to the precedent, the less effort is necessary to make the analogy. Put in economic terms, the cost of making an analogy increases as the distance between a new case and the closest precedent increases.

The appellate court observes whether the trial court held the defendant liable or not liable. It does not, however, observe the location of the case. This assumption reflects the fact that it is the trial court, and not the appellate court, that hears the evidence (James et al., 1992). It takes effort for the appellate court to discover where along the spectrum the case falls. If, say, the trial court finds the defendant not liable, the appellate court doesn't know whether the

[6] Some might construe the term legalist to imply that cases have a "right" answer or that judges should avoid any reference to ideology. We do not use the term in this way. Rather, we suggest that reasoning to the closest available precedent is a way for judges to economize on the effort required to resolve a case.

seriousness of the case was 0, .1, .35, or .5. Only a close examination of the trial court record will reveal this information. The appellate court must decide when it will expend its scarce resources to learn where the case lies.

First consider a one-period interaction between the trial court and appellate court. What happens when the appellate court observes a trial court disposition that favors the plaintiff (the defendant is liable)? In this situation, the appellate court suffers no loss because that outcome is consistent with its preferences. Accordingly, it spends the minimum possible effort affirming these decisions. For example, the appellate court might issue a one-line order or a cursory, unpublished decision. What about decisions where the trial court determines that the defendant is not liable? Assume that the appellate court suffers a loss of 100 if the case goes against its preference, which occurs when it is either .35 or .5. Given the no liability disposition, the appellate court can infer that each of these outcomes occurs 1/4 of the time. Thus, the appellate court will spend up to 50 (((1/2)*(100)) reviewing trial court outcomes that find no liability. If the appellate audit costs more than 50, the appellate court will simply affirm and suffer the loss. If it costs less, she will spend resources reviewing the record closely. If, through its investigation, the appellate court uncovers cases .35 or .5, it reverses. If it uncovers case 0 or .1, the appellate court affirms, as those cases are decided in a way the appellate court prefers.

This one-period example replicates the error correction models in the literature (Kornhauser (1992b); Shavell (1995)). The appellate court audits the trial court. If the appellate court finds a case in the undefined area that goes against its preferences, it reverses. Notably, there is no reason for the appellate court to develop law in this one-period interaction. The appellate court does not write an opinion because doing so would be costly (De Mesquita and Stephenson, 2002) and it would provide no future benefit given that the interaction lasts for one period.

Moving to a two-period interaction changes matters. This longer term perspective allows us to understand how developing the law might serve the long term interests of the appellate court. In period 1, suppose that the appellate court reviews a case where the trial court held the defendant liable. In so doing, the appellate court uncovers case .7. The appellate court affirms and declares that, in all future cases, the plaintiff prevails in all cases .7 and higher. What happens in period 2?

Let's say that the trial court draws case .5 in period 2. Given the new precedent promulgated in period 1, the closest analogous precedent is the .7

one. That precedent says the the plaintiff wins: the defendant is liable. The establishment of new precedent alters the way the trial court resolves case .5. Unlike the one-period model, the trial court now resolves cases 0, .1, and .35 in favor of the defendant. It resolves cases .5, .7, .95, and 1 in favor of the plaintiff. Recall the appellate court wants the defendant to be liable in cases .35, .5, .7, .95 and 1. By establishing new precedent in the first period, the appellate court gets closer to its desired resolution of the cases in the second period. Now, the appellate court is only willing to spend 33 (1/3*100) auditing not liable dispositions. If the cost of auditing exceeds 33, the appellate court will let the disposition stand with no review. If the cost of auditing is less than 33, she will review. In the one period model, the appellate court suffers a loss of, at most, of 50. In the two-period model, the appellate court suffers a loss of, at most, 33.

By establishing precedent, the appellate court provides a new case from which the trial court can draw analogies and, in so doing, alters the way trial courts decide future cases. The prospect of being able to establish precedent, then, feeds back into the appellate court's review of trial court dispositions in period 1. Unlike the one-period model, the appellate court spends resources auditing dispositions with which it agrees. It might affirm and write an opinion about what counts as too little precaution. The appellate court isn't just trying to correct trial court errors; it wants to develop the law in a way that aids the trial courts in the future.

Consider two examples of this practice. In Nightingale Home HealthCare, Inc. v. Anodyne Therapy, LLC[7], the trial court awarded the defendant attorney's fees in a trademark case, which the Lanham Act permits in exceptional cases. In an opinion by Judge Posner, the appellate court affirmed. An express goal of the opinion was to "clarify" when attorneys' fees should be awarded in trademark cases. Judge Posner rooted the test in abuse of process. He then applied the test to the facts, saying that the plaintiff had abused the process because it filed the claim in an attempt to "coerce a price reduction" from the defendant.

Miller v. Othello Packers, Inc.[8] involved a contract between a crop grower and a crop processor. The parties left the price open, to be determined by "tonnage and grading as the beans went through the processor's plant." The trial court awarded the grower a reasonable value of the crop. The Supreme

[7] 626 F.3d 958 (2010).
[8] 410 P.2d 33 (1966)

Court of Washington affirmed. It explained that the harvester failed to act in good faith because it had "left three truckloads of bean vines in the grower's fields." In so doing, the harvester reduced the value of the harvested crop and, with it, the price he would have to pay on the contract.

In both cases, the appellate court took time to write an opinion affirming the trial court. The opinions establish markers for future trial courts to follow. Before Nightingale a trial court may have found itself at sea when it tried to determine whether a set of facts qualified as exceptional circumstances. With little precedent it would take significant effort to justify an award of attorneys' fees. In this situation, the trial court may believe that the proper course is to analogize to the more common cases where fees are not shifted. But after Nightingale, trial courts understand that using trademark litigation to extract price concessions counts as abuse of process and, as a result, merits shifting fees. The next case might present actions close – but not identical to – a price concession. In light of Nightingale, the trial court will feel comfortable shifting fees in that future case; she can cite the Nightingale precedent and reason by analogy. Prior to Nightingale, this option was not available. Similarly, Miller tells courts that allowing a product to rot amounts to bad faith in a contract whose price based on the value of the processed seeds. Even though a subsequent case might involve a substantially different product–say delayed delivery of components for a new model of smartphone–the ability to analogize to the time-sensitive nature of the bean vines in Miller makes it easier for a future trial court to resolve the case.

These cases, we submit, are not atypical. Indeed, studies show that a significant number of published opinions affirm the result below (Mead, 2001). Neither Nightingale nor Miller involve the mistaken trial court that is the focus of the enforcement literature. In our model, the reason for these cases is readily apparent: by making law, the appellate court makes it easier (cheaper) for trial courts to decide future cases.

Three other insights follow from the observation that there is a benefit to developing precedent. First, the appellate court will spend more resources reviewing decisions it expects to reverse than decisions it expects to affirm. In this example, in period 1, the appellate court expects to reverse half the cases where the trial court finds the defendant not liable. Reversal carries a large potential for precedent building. To see why go back to period 1, where the trial court found the defendant not liable in cases 0, .1, .35, and .5.

Suppose that the appellate court audits a not liable disposition and uncovers

case .5. The appellate court reverses and says that defendants should be found liable in this case. Now, in period 2, the trial court will "flip" the disposition of case .35 and case .5 from not liable to liable. The precedent is directly on point for any subsequent .5 case. Case .35 is closer to the liable precedent (resting at .5) than to the not liable precedent (resting at .1). If instead the appellate court audits a first period disposition finding liability, it knows that the case lies at .7. .95 or 1. The best the appellate court can hope for is to establish the .7 precedent. In so doing, the appellate court changes the trial court's disposition of the .5 case in the next period. Notice that the appellate court can never use an affirmance of a liable disposition by the trial court in period 1 to alter how trial court in period 2 resolves case .35. The closest liable disposition, .7, is too far away from the .35 case to provide useful precedent to the trial court.

Put yourself in the mind of the appellate court. It knows that reversing a not liable disposition has the potential to change the trial court disposition in two kinds of cases in the future (.5 and .35). It also knows that affirming a liable disposition has the potential to change the trial court disposition in one kind of case in the future (the .5 case). Because two cases going its way are better than one, the appellate court is willing to spend more resources auditing decisions where it expects to reverse. The desire to create effective precedent leads to asymmetric appellate scrutiny: the appellate court looks harder at cases that have a chance of being reversed.

Turning to our next insight, consider dicta. Most commentary on the practice of dicta focuses on definitional issues (Abramowicz and Stearns, 2004). We take the hornbook definition as given – dicta are statements unnecessary for the resolution of the case. We ask why a court might issue dicta in the first place.

As scholars emphasize, reasoning by analogy relies on a comparison of the details in concrete controversies (Holmes (1870); Sunstein (1993)). With this guidance in mind, imagine a .7 case where the appellate court announces that all cases .35 and above should result in liability for the defendant. Because the statements about a hypothetical .35 case are far and unconnected from the facts of the actual .7 case, the dicta might not be that helpful to the trial court. Put another way, the trial court cannot easily analogize an actual .35 case to what the appellate court has said about a hypothetical .35 case. To provide effective guidance, dicta needs to be related to the actual case before the appellate court. The statements need to be close, but not too close to the holding. Statements too close to the holding fail to maximize the potential benefits of dicta for aiding the lower court's analogical reasoning. Statements too far away are ignored.

The appellate court must balance these two concerns.

Our model predicts that appellate courts will issue dicta and lower courts will make use of this guidance. The dicta statements won't involve an assertion of what the appellate judge prefers in all cases on a topic. Instead, they will be statements about what the appellate judge prefers in some closely related cases. Consider Myers v. Loudon County Public Schools[9], a constitutional challenge to the Pledge of Allegiance in the Fourth Circuit. Even though the Supreme Court had recently declined to decide the issue in Elk Grove v. Newdow,[10] the Myers majority upheld the Pledge on the basis of dicta in similar Supreme Court cases. In another instance of this practice, an Oregon Court of Appeals had to evaluate whether the DMV accorded due process when it suspended the plaintiffs' driver's license.[11] The court expressly relied on dicta from the Oregon Supreme Court to reverse the suspension. More systematic studies of dicta suggest that these cases are not anomalies. Klein and Devins (2013) code a sample of appellate and trial court opinions for the use of dicta. They find that, when available, lower courts overwhelmingly follow dicta, which is consistent with the predictions of our model.

3 The Model

A long-lived appellate court faces a series of short-lived trial courts. The assumption of short-lived trial courts is consistent with a judicial system involving many trial courts and a single appellate court. With many trial courts, free riding will be prevalent, making coordination over actions which impact future payoffs to other trial courts hard to do. Each trial court thus defaults to maximizing its own one-period payoff. Time is discrete, indexed by $t = 0, 1, 2$.

3.1 Cases

Each period, a trial court draws a case $x \in [0, 1]$ from the uniform distribution. A case is a set of facts that describes an activity. The larger x is, the higher the social costs and the lower the social benefits of the activity. At time 0, an initial precedent is exogenously established. This precedent instructs that: (1) no liability should arise in cases below $\underline{x} < \frac{1}{2}$; and (2) liability should result

[9] 418 F.3d 395 (4th Cir. 2005).
[10] 542 U.S. 1 (2004).
[11] Hays v. DMV, 216 P.3d 902 (Or. Ct. App. 2009).

for cases above $\overline{x} = 1 - \underline{x}$. The precedent might come from an (unmodeled) Supreme Court or from the appellate court itself. At time 0, the interval of uncertain law is $[\underline{x}, \overline{x}]$. At time 0, the appellate court has not spoken about which activities beyond the initial precedent should result in liability. In terms of preferences, assume that the appellate court does not wish to expand the no liability precedent. It wants to restrict the holding from the Supreme Court or from its own prior case law to its facts. One can thus think of $\underline{x}$ as the appellate court's preferred cutoff, slicing the case space into activities that result in no liability and activities that result in liability.[12]Over time, the appellate court will set new law about when a defendant should be found liable. Define the lowest such activity as of time t as $\widetilde{x}_t$. Denote the set of possible cases as X and the set of possible precedents as $\underline{X}$ and $\widetilde{X}$.

3.2 Trial Courts

Trial courts fall into two categories: $i \in \{1, 2\}$. The category 1 trial court judge is legalist and occurs in the population with probability β_1. In a way we will formalize momentarily, this category of trial court judge wishes to minimize the effort she must expend to analogize a new case to appellate precedent. The category 2 trial court is a realist. She wants to expand the applicability of the initial "no liability" precedent beyond its present position. The realist judge occurs in the population of trial court judges with probability $\beta_2 = 1 - \beta_1$. We assume that the appellate court observes the category of trial court hearing the case and can tailor its auditing decision accordingly.

3.2.1 Preferences and Strategy: The Legalist Trial Court

After receiving a case, the legalist trial court must render a decision and justify it.[13] Justification requires effort. The effort required to find the defendant not liable depends on how analogous the case is to the not liable precedent case. The closer the case is, the less effort will be required to justify ruling

[12]We study the case where the preferred cutoff can be found in the inherited prior case law. The issue is that the law is undefined in the area above the cutoff. That is to say, the trial courts do not know whether to limit that precedent or face large analogical reasoning costs in doing so. The appellate court provides guidance by issuing new decisions. Baker and Mezzetti (2012) study the case where the preferred cutoff is unknown to the appellate court.

[13]Kim (2007) notes that "[W]hen deciding cases, judges do at least two things. They determine the outcome of the dispute before them, and they offer reasons for their decision that connect the facts to applicable legal doctrine." The latter step requires effort and is what we model here.

for the defendant.[14] Formally, we model the effort cost incurred in finding the defendant not liable as $x - \underline{x}$. Likewise – and for similar reasons– the effort cost associated with finding the defendant liable is $\widetilde{x}_t - x$.[15] Figure 2 depicts these costs.

The legalist trial court pure strategy is a decision – liable, not liable – for each possible case. It makes this decision in light of the existing precedent bounds. Formally, the legalist trial court's pure strategy is function mapping the set of possible cases and precedents into a decision: $s_{LTC} : X \times \underleftarrow{X} \times \widetilde{X} \to \{L, NL\}$. The trial court makes this decision to minimize its effort costs.

Now take the legalist trial court hearing a case $x \in [\underline{x}, \widetilde{x}_t]$. Suppose the trial court finds the defendant not liable. For this disposition, the reasoning or effort cost is $x - \underline{x}$. If instead the trial court finds the defendant liable, rationalizing the finding costs $\widetilde{x}_t - x$. Given these costs, the decision rule of the legalist trial court is simple:

$$x < \frac{\underline{x} + \widetilde{x}_t}{2} = x_t^* \qquad \text{Not Liable} \tag{1}$$
$$x \geq \frac{\underline{x} + \widetilde{x}_t}{2} = x_t^* \qquad \text{Liable}$$

Distance from the precedent border determines the legalist trial court's decision. If the case lies closer to the not liable precedent than the liable precedent, the trial court finds the defendant not liable. Otherwise, it finds the defendant liable. This, we submit, is one way to model a effort-motivated or legalist judge.[16] She just cares about which appellate precedent is closer to the case at hand because

[14] A skeptic might respond that a trial court can always exert no effort whatsoever. Perhaps the trial court could flip a coin to see determine which side prevails. Doing so would, however, violate the Federal Rules of Civil Procedure. In a bench trial, Rule 52 requires that "the court must find the facts specially and state its conclusions of law separately." In awarding summary judgment, the trial court must "state on the record the reasons for granting or denying the motion.for summary judgment." We assume that trial courts look to appellate court precedent to justify dispositions.

[15] We have bench decisions in mind in the model. But we might also think that the trial court in instructing the jury might emphasize one set of precedents over another.

[16] The same model can accommodate a trial court whose objective is to follow the appellate court precedent. Suppose the trial court suffers a loss if he decides the case incorrectly from the appellate court's point of view. The trial court does not know the appellate court's cutoff point. It must infer the cutoff from the precedent interval $[\underline{x}, \widetilde{x}_t]$. Given this interval, the trial court treats the appellate court's cutoff as a random variable – call it x^C – distributed uniformly on $[\underline{x}, \widetilde{x}_t]$. After drawing a case x, suppose the trial court finds the defendant not liable. The probability he believes this decision is mistaken is $prob\{x^C < x\} = \frac{x - \underline{x}}{\widetilde{x}_t - \underline{x}}$. Suppose instead he finds the defendant liable. The probability he believes this decision is mistaken is $prob\{x^C > x\} = \frac{\widetilde{x}_t - x}{\widetilde{x}_t - \underline{x}}$. Setting equal the two expected losses gives the same decision rule, x_t^*, as in the text.

close analogies are cheaper to do in terms of the costs of legal reasoning. The cutoff is the average value of the precedent in the region of uncertain law. The legalist trial court doesn't care whether the plaintiff or the defendant wins; just about her effort. Thus, we assume they are indifferent to reversal.[17]

3.2.2 Preferences and Strategy: The Realist Trial Court

Recall that the realist trial court wishes to expand the applicability of the initial "no liability" precedent beyond its present position. Define as l the loss the realist trial court suffers if she ever finds a defendant liable. Given that that appellate court prefers to find the defendant not liable in many cases (all those above $\underline{x}$), the preferences of the realist trial court and the appellate court conflict. Like the legalist trial court, the realist trial court's can either find the defendant liable or not liable. It does so in light of the probability of the audit by the appellate court. Denote by p_2 the probability of a successful appellate court audit of a not liable disposition by the realist trial court. The set of probabilities is P_2. The realist trial court's strategy takes every case and every audit probability and gives an outcome: liable, not liable. Formally, it is a function $s_{RTC} : X \times P_2 \rightarrow \{L, NL\}$. For reasons described in a moment, the appellate court will never be able to audit perfectly (i.e., $p_2 < 1$). So, our set-up makes matters easy. The realist trial court has a dominant strategy: always find the defendant not liable. The expected loss to the realist trial court of finding the defendant liable is l. The expected loss from finding the defendant not liable is $p_2 l$ – a value always less than l.

3.3 The Appellate Court: One Period Model

The appellate court doesn't immediately know the true location of x. It can only observe the the category of the trial court – legalist or realist – and the outcome – liable or not liable.[18] The appellate court can, however, spend effort auditing the trial court's decision. It can examine the trial record in detail, for example. We model this decision as the appellate court selecting audit

[17] One might suspect that trial courts who don't care about outcomes might still care about reversal. Reversal signals they are bad judges. Reversal costs are sometimes modeled as resulting from a failure of a lower court judge to get his preferred outcome (McNollgast 1995). We follow this approach here.

[18] Our assumption is basically that the appellate court doesn't know how analogous the case in the trial court is to the appellate court precedent without some investigation. We model it as the appellate court's having an inability to observe the fact location without spending resources.

probabilities. Denote the probability of a successful audit of a no liability disposition rendered by trial court i as $p_i \in [0, 1]$. The probability of a successful audit of a liable finding is $q_i \in [0, 1]$.[19] Upon a successful audit, the appellate court learns the true location of x.

There is a cost to the appellate court of auditing. The cost is the same for liable and not liable dispositions. It is represented by $c(\cdot)$. The cost function satisfies the usual conditions, namely $c(0) = 0, c' > 0, c'' > 0; c'(0) = 0; c'(1) = \infty$. If the audit fails, the appellate court doesn't learn enough about the location of x to reverse the trial court (it cannot find reversible error). Thus, it lets the trial court decision stand.[20] As noted above, the appellate court prefers that all cases above $\underline{x}$ result in liability. In other words, $\underline{x}$ is the appellate court's cutpoint in the case space. Denote as λ the loss the appellate court suffers if the case is resolved in a way it disfavors.

The strategy of appellate court consists of an audit probability for each potential resolution of the case by each category trial court. The possibilities are: (1) a legalist trial court finds the defendant not liable; (2) a legalist trial court finds the defendant liable and (3) a realist trial court finds the defendant not liable. In terms of the formalities, denote as p_1 the audit probability applied to a legalist trial court finding the defendant not liable; q_1 as audit probability applied to a legalist trial court finding the defendant liable; and p_2 as the audit probabilty applied to a realist trial court finding the defendant not liable. Because the realist trial court always finds the defendant not liable–recall that this is the realist trial court's dominant strategy–we need not consider the scenario where it finds the defendant liable.

We assume that the appellate court selects and commits to the audit probabilities before nature resolves the uncertainty regarding the case location or the category of trial court to hear the case.[21] In the one-period model, the appel-

[19] We will alternatively refer to the audit probabilities as probabilities or appellate court effort per case. They are interchangeable, given that any level of effort corresponds to a level of audit success. For a similar approach to modeling the successful discovery of information, see Che and Karthnik (2009).

[20] The assumption that the appellate court must locate the case before reversing rules out the following course of events: The realist trial court finds the defendant not liable. The appellate court – knowing a realist trial court made the ruling – reverses summarily, without expending any effort. Anticipating immediate reversal if his type is found out, the realist trial court spends effort trying to convince the appellate court it was a legalist trial court instead. Rather than investigate this asymmetric information, we assume that the trial court's type is observable, but the case location is not. Then, in line with actual practice, we assume that, before reversing a trial court, the appellate court must be comfortable saying there was reversible error. Put another way, the appellate court must expend the effort and locate x in order to reverse.

[21] In the penultimate section of the paper, we discuss the likely implications of relaxing the

late court's sets its audit strategy to minimize its expected losses. Formally, its program is given by:

$$\max_{0\leq p_1\leq 1; 0\leq p_2\leq 1; 0\leq q_2\leq 1} \beta_1\{-(x^*(\underline{x},\overline{x})-\underline{x})(1-p_1)\lambda-\Pr[NL|\underline{x},\overline{x}]c(p_1)-\Pr[L|\underline{x},\overline{x}]c(q_1)\}$$
$$+\beta_2\{-\lambda(1-p_2)[1-\underline{x}]-c(p_2)\}$$

The first term is the appellate court's expected loss when the legalist trial court hears the case. It consists of three parts: (1) the probability that legalist trial court finds the defendant liable in a case above the appellate court's preferred cutoff[22] times the probability of a failed audit times the loss from a resolution the appellate court disfavors; (2) the cost of auditing the not liable disposition times the probability of a not liable disposition; and (3) the cost of auditing a liable disposition times the probability of a liable disposition. Notice that whether the legalist trial court finds the defendant liable (or not liable) depends on the precedent bounds. That is because following the closest appellate court precedent minimizes effort for the legalist trial court. The second term is the expected loss when the realist trial court hears the case. Such a trial court always finds the defendant not liable. As a result, this term has two parts: (1) the probability the case lies above the appellate court's cutoff times the probability of a failed audit times the loss from the incorrect resolution and (2) the cost of auditing not liable dispositions.

3.4 Timing

Although hinted at thus far, we can now in a position to be precise about the timing of the interaction between the appellate court and the trial court. It runs as follows:

1. The appellate court picks: (a) the probability of successfully auditing a not liable decision by each category of trial court $\{p_1, p_2\}$; and (b) the probability of successfully auditing a liable decision by the legalist trial court q_1.

2. Nature draws a case, x, and a category of trial court to hear the case.

assumption that the appellate court can commit to the audit strategy.

[22] Recall that the legalist trial court finds the defendant not liable in cases below x^*. The legalist court thus renders a not liable ruling inconsistent with the preferences of the appellate court if the case arises in the interval $[\underline{x}, x^*]$. The probability of this event is $x^* - \underline{x}$.

3. The legalist trial court decides the case according to its cutoff strategy, x_t^*. The realist trial court finds the defendant not liable.

4. The appellate court applies the audit probability suited for that trial court's resolution.

5. The appellate court reverses if it discovers that the trial court found the defendant not liable in a case above $\underline{x}$. The appellate court affirms if: (1) it discovers that the trial court found the defendant not liable in a case below $\underline{x}$; (2) it discovers that trial court found the defendant liable in a case above $\underline{x}$; or (3) the audit fails and the appellate court doesn't learn the location of x.

6. Payoffs are realized.

4 One-Period Benchmark

The results of the maximization program in expression (1) are intuitive. The appellate court never audits dispositions where the defendant is found liable. The appellate court suffers no loss from this outcome. So, it does not spend resources checking such a decision. Further, the appellate court gains nothing from using a review of liable disposition to set precedent. On the other hand, the appellate court always audits when the defendant is found not liable. It then reverses upon successful detection of a case above its preferred cutoff.[23]

Formally, we have

Proposition 1 *In a one-period model, the appellate court audits trial court decisions finding no liability with positive probability,* $p_1^{**} > 0$*;* $p_2^{**} > 0$*. It never audits liability dispositions by the legalist trial court* $q_1^{**} = 0$*.*

In the one-period model, the appellate court does not issue opinions. The practice of defining the law for application in the trial courts is absent. The one-period model amounts to error correction by the appellate court, with the inclusion of a cost of analogical reasoning in the utility function of the legalist trial court.

[23] We use a double-star superscript to denote the values of the appellate court audit strategy that solves the one-period program. That way, we retain the use of the single star for the dynamic model to come.

5 The Dynamic Optimization Problem

In the dynamic model, we ask whether the appellate court makes law and, if so, how. The legalist trial court's decision cutoff depends on the average value of cases where precedent provides no conclusive guidance. The bigger this average, the higher the cutoff is: the more cases where the trial court finds the defendant not liable instead of liable. If the appellate court has little precedent defining activities that result in liability, the trial court must work hard to justify holding the defendant liable. The analogous appellate precedent doesn't exist. Rather than expend effort reasoning from distant precedent, the trial court finds the defendant not liable instead. As noted in the numerical example, by making new law about what activities result in liability, the appellate court makes it easier for the trial court to find the defendant liable in closely analogous cases. The legalist trial court, as a result, will be more apt to find the defendant liable. This outcome, of course, is what the appellate court prefers. The appellate court audit of trial courts will be done with an eye toward these future benefits

In terms of the optimization, the state variable is $\widetilde{x}_t$ – the lowest activity where the appellate court has previously spoken and said the case should result in liability. The control variables are threefold. First we have p_{1t} and p_{2t} – the probability of successfully auditing a not liable finding of category 1 and 2 trial courts respectively. Second, there is q_{1t} – the probability of successfully auditing a liability finding by the legalist trial court. The choice of these variables depends on how much of the law the appellate court has defined previously – the location of $\widetilde{x}_t$.

Assuming a discount factor δ, the appellate court selects the control variables to maximize its discounted stream of payoffs. Denote the value function $V(\widetilde{x}_t)$. It can be expressed as

$$V(\widetilde{x}_t) = \max_{0\leq p_{1t}\leq 1; 0\leq q_{1t}\leq 1; 0\leq p_{2t}\leq 1} \beta_1\{-\lambda(1-p_{1t})[x_1^*-\underline{x}]-\Pr[NL]c(p_{1t})-\Pr[L]c(q_{1t})\} \\ + \beta_2\{-\lambda(1-p_{2t})[1-\underline{x}]-c(p_{2t})\} + \delta E_t V$$

In this expression, $E_t V$ is the appellate court's expected value function in the next period. This expected value function consists of a number of terms and

equals

$$E_tV = V(\widetilde{x}_t)\left([1-\widetilde{x}_t+\underline{x}]+\beta_1\{(1-p_{1t})\int_{\underline{x}}^{x_1^*}dx+(1-q_{1t})\int_{x_1^*}^{\widetilde{x}_t}dx\}+\beta_2(1-p_{2t})\int_{\underline{x}}^{\widetilde{x}_t}dx\}\right)$$
$$+\beta_1\left(p_{1t}\int_{\underline{x}}^{x_1^*}V(x)dx+q_{1t}\int_{x_1^*}^{\widetilde{x}_t}V(x)dx\right)+\beta_2 p_{2t}\int_{\underline{x}}^{\widetilde{x}_t}V(x)dx$$

The two regions of certain law are $[0,\underline{x}]$.and $[\widetilde{x}_t,1]$. In our numerical example, these two sets were {0,.1} and {.95,1}. If existing precedent is on point with the appealed case, the appellate court cannot use the case to make new law. In the model (as opposed to numerical example), the appellate court also cannot make new law if the audit fails to locate x. The first term adds these two probabilities together and multiplies them by $V(\widetilde{x}_t)$. The reason: The environment doesn't change if the appellate court cannot make new law. It thus faces the same choices in period $t+1$ that it did in period t.

The second term is the expected payoff from successfully auditing the legalist trial court. The interval of uncertain law is separated into cases where the defendant is found not liable and cases where the defendant is found liable. If the audit is successful, the state of the appellate court case law shifts in next period. The amount of the change depends on the location of the case on the unit interval. That is why the value function is integrated over the case space in the second term. The third term is expected payoff from successfully auditing the realist trial court.

Lemma 1 in the appendix shows that, if the appellate court's loss from erroneous decisions is large, a unique and differentiable value function, $V(\widetilde{x}_t)$, exists. It decreases in $\widetilde{x}_t$ and reaches it maximum at $\widetilde{x}_t = \underline{x}$ – when the appellate court has defined the law for all possible cases.

The next proposition sets out how the appellate court makes law in this dynamic setting.

Proposition 2 *(a) In the dynamic model, the appellate court audits trial court decisions whose resolution do not result in a loss to the appellate court (i.e., $q_{1t}^* > 0$). (b) The audit of the legalist trial court is asymmetric ($p_{1t}^* > q_{1t}^*$).*

Proposition 2 helps to explain a number of institutional features of the judiciary. Depending on the location of the case and the success of the audit, the

appellate court might (1) affirm and make law; (2) affirm with minimal effort (i.e. through a cursory, unpublished opinion)[24]; or (3) reverse and make new law. Specifically, if the case lies in the range of settled law, the appellate court affirms via unpublished opinion. Likewise, if the audit fails–meaning that the appellate court cannot ascertain the location of x with sufficient confidence–the appellate court affirms the trial court disposition in an unpublished decision. If the audit of a not liable disposition succeeds and the case lies in the region of uncertainty, the appellate court reverses and makes new law.[25] If the audit of a liable finding succeeds and the case lies in the region of uncertainty, the appellate court affirms and makes new law.

Part (a) of proposition 2 establishes that the appellate court audits cases whose disposition it knows it will agree with. The appellate court wants to make law. Liable findings offer a potential opportunity to do so. Indeed, the appellate court might hold oral argument or ask for additional briefing, all in a case where it knows an affirmance is in the offing.

Part (b) of the proposition predicts that the appellate court will scrutinize "not liable" dispositions more intensely than "liable" dispositions. In other words, the appellate court's auditing resources will tilt toward review of cases where the appellate court might disagree with the trial court's disposition. Reversal in such cases is especially useful for shifting down the liability precedent. Indeed, the best case for setting law from the appellate court's perspective rests a smidge above $\underline{x}$. The appellate court wants to take this case, reverse and instruct that all cases above $\underline{x}$ result in liability. That way, the appellate court can make the most law in a single decision, thereby minimizing the extent – and cost – of future audits.

In light of this analysis, part (b) of this proposition yields the first testable prediction from the model.

Prediction 1: Appellate courts should be more likely to make new law through reversals of trial courts than through affirmances.

Next we consider how appellate scrutiny of trial court dispositions changes as the law becomes more defined.

[24] For convenience we will refer to these cases as unpublished cases.

[25] There is some evidence that appellate court judges act strategically when it comes to the decision whether to publish a case (Law, 2004). Our model of a single long-lived appellate court means that these strategic elements do not come into play. We leave the modeling of publication decisions on multi-member courts to future research.

Proposition 3 *If $\lambda > c'(p^*_{1t})$, appellate scrutiny of not liable dispositions decreases as the law becomes more defined; that is, $\frac{\partial p^*_{2t}}{\partial \widetilde{x}_t} > 0$ and $\frac{\partial p^*_{1t}}{\partial \widetilde{x}_t} > 0$.*

As the law gets more defined, it becomes less likely that an audit of a not liable finding will uncover a case in the range of uncertain law. Instead, a successful audit is likely to reveal a case overlapping with an existing appellate precedent. Given the overlap, the appellate court can't use the case to create new law. Since the benefit of the audit falls as the interval of uncertain law shrinks, the appellate court spends less auditing these dispositions of the trial courts.[26]

Proposition 3 suggests that appellate courts will invest more heavily in reviewing trial court dispositions when the law is its infancy, perhaps shortly after a statute has passed or just after the Supreme Court issues a decision. In each scenario, the appellate court will be searching for cases decided in the trial courts to make law. Given the lack of precedent, this search will often be fruitful. This leads to the next prediction from the model.

Prediction 2: Over time, the appellate court will publish fewer and fewer decisions on a specific topic; that is, the ratio of unpublished to published decisions will rise.

The available evidence is consistent with this prediction. One prominent example is petitions for habeas corpus. The law in this area is fairly well defined and, consequently, it is unlikely to be an effective use of appellate court resources to publish many decisions in this area. Though systematic studies of the degree to which court fill in habeas law do not appear to exist, empirical evidence shows that the ratio of unpublished to published opinions for habeas appeals is substantially higher relative to other types of appeals (Mead, 2001). There are other areas of law where auditing will not be as intense. For example, one judge has suggested appeals of immigration cases and Social Security cases tend not to implicate the "law-declaring function" of appellate courts because most issues are determined by precedent (Jones, 1995).

[26] The condition that $\lambda > c'(p^*_{1t})$ ensures that $\frac{\partial p^*_{1t}}{\partial \widetilde{x}_t} > 0$. Intuitively, the condition means that, as the law becomes less well-defined, the marginal benefit of an audit of a not liable increases faster than the marginal cost. As to the audit of liable findings, the comparative statics are ambiguous; they can't be signed without knowing more about the aspects of the function V. We can say that, as the law converges, the audit of a liable finding by a legalist court goes to zero.

This prediction suggests another likely pattern. After there is an exogenous shock to precedent–such as a Supreme Court opinion that reverses longstanding law–appellate courts should publish more opinions. The exogenous change can create a new area of undefined law and it will be in the interest of appellate courts to provide guidance in that interval. Examples of this phenomenon include the need to define what a substantive reasonable sentence is in the wake of Gall v. United States[27] and the development of a standard of review for Second Amendment claims after District of Columbia v. Heller[28] and McDonald v. City of Chicago[29] (Levy, 2013).

To close this section, we emphasize that the appellate court does strictly better as it defines the law. Its overall losses – given the optimal review strategy – shrink as the interval of uncertain law collapses.

5.1 Convergence of Law

The appellate court can put in little effort and hope it successfully uncovers a case in the interval of uncertainty (this follows since $c'(0) = 0$). Given this assumption, the appellate court audits until the interval of uncertain law vanishes ($\widetilde{x}_t = \underline{x}$). If the law spans the entire spectrum, the legalist trial court will always find an appellate court case on point and follow it. The legalist trial court need not devote any effort to reasoning from an analogous case. The appellate court understands as much. It thus affirms the legalist court's disposition via unpublished decision.

The realist trial court is a different story. Given this trial court's preferences over outcomes, this court will continue to find the defendant not liable. After the law converges, then, the steady state involves (1) the realist trial court finding the defendant not liable in the face of a liable precedent directly on point and (2) the appellate court spending resources auditing not liable dispositions from the realist trial court.

We predict, then, asymmetric treatment of trial courts in the limit. The realist trial court faces appellate scrutiny. The legalist trial court does not. We have the following two results.

Proposition 4 *(A) Eventually, the law converges – a precedent is defined for every possible case. (B) After convergence, the appellate court no longer audits*

[27] 552 U.S. 38 (2007).
[28] 554 U.S. 570 (2008).
[29] 130 S.Ct. 3020(2010).

dispositions by legalist trial courts; it continues to audit not liable dispositions by realist trial courts.

After the law converges, the appellate court's long run loss is given by

$$V(\underline{x}) = \frac{-\beta_2\{(1-\underline{x})\lambda(1-p_2^{**}) - c(p_2^{**}))\}}{1-\delta}$$

Realist trial courts do not respond to precedent and appellate review is not perfect. As a result, the appellate court suffers an expected loss anytime a realist trial court hears a case. Some cases where the appellate court would have preferred liability go undetected and are not reversed.

Note that, in the long run, the appellate court suffers losses even though it has fully defined the law. This outcome arises because (1) the initial decision is delegated to an agent who holds different, conflicting preferences and (2) appellate review is costly. The long run losses are the agency costs of having a judicial hierachy. The appellate court can control the ealist trial court if it devotes enough resources to detect each time they deviate from established precedent. Yet devoting so many auditing resources to detection is not cost-justifed.

The appellate court's per period loss is less in the long run than in the one-period model. By making law, the appellate court enables the legalist trial court to costlessly follow its precedent. It provides cases on point, eliminating the need for the legalist trial court to do analogical reasoning. In so doing, the appellate court eliminates all its losses from delegating decisionmaking to the legalist trial court.

A testable prediction flows from this proposition.

Prediction 3: Appellate courts will determine the amount of resources for review based on the type of trial court rendering the disposition. The appellate court will more heavily audit trial courts who hold conflicting policy preferences.

Measuring the resources that judges devote to the review of lower court opinions is, of course, a difficult task. But a number of empirical studies show that the behavior of lower courts appears to differ in a way that depends on the policy preferences of reviewing courts. These findings suggest that lower courts respond to the threat of more searching review by appellate courts that have conflicting policy preferences. Schanzenback and Tiller(2008) show this result in the sentencing context. They find that federal district court judges are

more likely to depart from the federal sentencing guidelines when the reviewing court is politically aligned with the reviewing court. This evidence suggests that appellate courts are likely to expend more effort reviewing a sentence from a lower court with conflicting policy preferences and, moreover, lower courts appear to be responding to that preference. Another study shows that the degree of political conflict with reviewing courts appears to affect whether federal district courts choose to publish their decisions (Choi et al., 2012). This finding suggests that, when the stakes are high, policy conflicts will lead appellate judges to scrutinize cases more closely. Lower courts respond to this threat by choosing not to publish cases that the appellate court is likely to reverse. While these studies do not provide a direct measure of audit intensity, they permit inferences that are consistent with our model.

6 Holding and Dicta

Thus far, we have assumed that the promulgation of law took a specific form. If the appellate court uncovered a case located at, say, .7, it announced that all cases above .7 should be decided against the defendant. With this case as a vehicle, the appellate court could not make law about cases below .7. The appellate court's opinion was limited by the facts as presented the trial court. This section relaxes this assumption. If the appellate court reviews a case at .7, it can say in the opinion that all cases above, say, .5 should result in liability. The opinion, then, consists of two parts: (1) a holding – all cases .7 and above should result in liability and (2) dicta – cases in the interval [.5,.7] should result in liability.[30]

In our example, any statements about case .5 are unnecessary for holding the defendant liable who has a .7 level of activity.

To study the impact of dicta, suppose that, if the the appellate court uncovers a case in the interval of uncertain law, it can issue dicta, $\Delta \in [\underline{x}, \widetilde{x}_t]$. In finding the defendant liable, the legalist trial court can then justify its decision by reference to the prior holding or dicta, or a combination of the two Suppose that this trial court's effort cost of holding the defendant liable is given by

$$[\alpha\Delta + (1-\alpha)\widetilde{x}_t] - x$$

[30] Black's Law Dictionary (2009) defines "obiter dictum" as "[a] judicial comment made while delivering a judicial opinion, but one that is unnecessary to the decision in the case and therefore not precedential (although it may be considered persuasive)." We follow this definition here.

In this expression, α is the weight the trial court allocates to dicta and $(1-\alpha)$ is the weight allocated to the holding. According to legal scholars, the further away the dicta is from the holding, the less likely it is to be persuasive (Dorf, 1994). In other words, the distance between dicta and the holding determines its influence. To capture this effect in the simplest way, suppose that

$$\alpha = \frac{\Delta - \underline{x}}{\widetilde{x}_t - \underline{x}}$$

If the dicta spans the entire interval of uncertain law, the trial court places no weight on it. The closer the dicta is to the holding, the more weight the trial court places on the statements. Suppose the appellate court aggressively issues dicta in a case involving liability – setting Δ close to the no liability precedent border. With this move, the appellate court maximizes the chance that the trial court will ignore the dicta and rely solely on the holding. In other words, the trial court won't be able to put the dicta into context and thus it won't help reduce its cost in justifying a finding of liability. As the dicta gets closer and closer to the holding, it becomes more likely that the trial court will rely on the dicta. At the same time, the dicta doesn't span much of the interval – it doesn't say much beyond the holding. As a result, the dicta doesn't lower the costs of finding the defendant liable by all that much. As we will see, the appellate court extends the dicta until the increase in the chance it fails to persuade the trial court just offsets the increase in liability findings if, in fact, it does persuade.

With this understanding of dicta in place, the trial court's dispositional cutoff becomes

$$x < \frac{\underline{x} + [\alpha\Delta + (1-\alpha)\widetilde{x}_t]}{2} = x_t^* \qquad \text{Not Liable}$$
$$x \geq \frac{\underline{x} + [\alpha\Delta + (1-\alpha)\widetilde{x}_t]}{2} = x_t^* \qquad \text{Liable}$$

The appellate court can (and will) issue a different amount of dicta for each case in the interval of uncertainty. Technically, the ability to issue dicta changes the state variable in the optimization problem. Before the state variable was the lowest activity where the appellate court had previously found the defendant liable. Now the state variable is the dicta/holding combination associated with that activity, $\widetilde{x}_t$. Define this new state variable as

$$\widehat{x}_t = \alpha\Delta(\widetilde{x}_t) + (1-\alpha)\widetilde{x}_t$$

The timing of the interaction each period runs like this:

1. The appellate court selects an audit review strategy $\{p_1, q_1, p_2\}$ and its plan for dicta, $\Delta(x)$, for each $x \in [\underline{x}, \widetilde{x}_t]$

2. A case and trial court is drawn. The trial court renders a decision: not liable or liable.

3. The appellate court applies the audit strategy associated with that disposition and category of trial court. If it discovers a case in the interval of uncertainty, the appellate court issues new dicta ($\Delta(x)$) and a new holding $\widetilde{x}_t$. In that case, the state variable changes accordingly.

Given a state variable $\widehat{x}_t$, the appellate court selects $\{p_{1t}\ q_{1t}, p_{2t}\}$ and $\Delta_t(\cdot)$ to maximize

$$V^{\Delta}(\widehat{x}_t) = \max_{0 \le p_{1t} \le 1; 0 \le p_{2t} \le 1; 0 \le p_{1t} \le 1;\ \{\Delta_t(\cdot) \in [\underline{x}, \widetilde{x}_t]\}} \beta_1\{-\lambda(1-p_{1t})[x_t^* - \underline{x}] - x_t^* c(p_{1t}) - (1-x_t^*)c(q_{1t})\}$$
$$+ \beta_2\{-\lambda(1-p_{2t})[1-\underline{x}] - c(p_{2t})\} + \delta E_t V^{\Delta}$$

In light of the possibility of dicta, the value function expected in the next period can be written as

$$E_t V^{\Delta} = V(\widehat{x}_t)\left([1-\widetilde{x}+\underline{x}] + \beta_1\{(1-p_{1t})\int_{\underline{x}}^{x_t^*} dx + (1-q_{1t})\int_{x_t^*}^{\widetilde{x}_t} dx\} + \beta_2(1-p_{2t})\int_{\underline{x}}^{\widetilde{x}_t} dx\}\right)$$
$$+ \beta_1\left(p_{1t}\int_{\underline{x}}^{x_t^*} V^{\Delta}(\widehat{x}(x,\Delta)dx + q_{1t}\int_{x_t^*}^{\widetilde{x}_t} V^{\Delta}(\widehat{x}(x,\Delta)dx\right) + \beta_2 p_{2t}\int_{\underline{x}}^{\widetilde{x}_t} V^{\Delta}(\widehat{x}(x,\Delta))dx$$

There are two differences from the baseline model. First, the cutoff for the legalist trial court is lower. Second, if the appellate court successfully audits a case in the interval of uncertain law, it can issue a holding and dicta. Thus, in the second line the value function (the expected value from a successful audit) depends both on the case draw (x) and the amount of dicta Δ.

As noted, the benefit of dicta is that it can lower the trial court's decision costs associated with finding the defendant liable. The trial court, however, ignores appellate court statements too far removed from the holding. Each period, the appellate court balances these two concerns and issues dicta optimally. Formally, we have

Proposition 5 *(A) The appellate court always issues dicta (i.e.,*$\Delta^*(x) = \frac{x+\underline{x}}{2}$*); (B) The appellate court is strictly better off when it can issue dicta than when it cannot; (C) The legalist trial court is strictly better off when the appellate court issues dicta.*

The literature on dicta provides support for the predictions of our model. While scholars and judges debate the extent to which judicial opinions should go beyond the specific facts at issue, there is no question that they regularly do so. Schauer (1995) goes so far as to argue that the presence of dicta is inevitable if judges have to give reasons to support their rulings. As he puts it, "every time a court gives a reason it is, in effect, giving an advisory opinion." If this proposition is correct, that means that, to some degree, an appellate court issues dicta whenever it writes an opinion. This near-universal use of the practice is what the model predicts.

The commentary also suggests that dicta is useful to lower courts. As Judge Pierre Leval has explained, dicta "can assist future courts to reach sensible, well-reasoned results" (2006).[31] Some scholars emphasize that dicta can have the beneficial effect of providing guidance in future cases (Katyal, 1998) and others argue that broad statements are consistent with the need for judges to settle disputes in an authoritative manner (Alexander and Schauer, 1997). And, as noted earlier, empirical studies of dicta confirm that lower courts often use the guidance that dicta provides (Klein and Devins, 2013). These findings are consistent with the model's predictions.

7 Comments on Modeling Assumptions

7.1 Appellate Court Can Commit to Audit Intensity

We assume that the appellate court can commit to an audit strategy before nature resolves (1) the uncertainty about the location of the case; and (2) which kind of trial court will hear the case. The commitment power, we suspect, is not needed for the results. In our set-up, the trial courts have dominant strategies: strategies independent of the audit intensity the appellate court selects. Thus, the appellate court doesn't benefit from increasing the audit intensity, changing trial court behavior and, then, not following through to save on auditing resources. In other words, our setup avoids the enforcer's dilemma.

[31] While Judge Leval does believe that dicta has this benefit, he argues that courts engage in the practice too broadly and in a manner that exceeds their appropriate authority.

Of course, a more detailed model might have, say, the realist trial court also care about both a reputational costs from reversal and the loss from not getting their preferred outcome. With this framework, the appellate court would want to threaten a high audit probability to induce compliance by the realist trial court and then not actually audit ex post to save resources. Knowing this, the realist trial court wouldn't comply. The equilibrium, we suspect, would likely involve some mixing by the appellate court (on this sort of equilibrium, see Cameron et al. (2000)). While interesting, our focus is more on precedent building than appellate court commitment. Indeed, given that the appellate court is a long run player here, it would likely have an incentive to build a reputation for commitment.

7.2 Cases Don't Settle

Throughout our analysis, we assume that cases don't settle, which is an unrealistic assumption. Suppose otherwise. Specifically, assume that parties settle cases before the legalist trial court where the law is clear (that is, $x \notin [\underline{x},\widetilde{x}]$). The issue is what effect, if any, will this settlement have on the review strategy of the appellate court of cases that don't settle. It is easy to that settlement improves the yield rate on the audit of the dispositions by the legalist trial court. Given settlement, the appellate court knows that any case that goes to legalist trial court and are appealed must involve new issues, issues where the law is uncertain. These cases present the opportunity to make law and, as a result, alter the future decisions in the legalist trial courts. With settlement, the appellate court will never audit successfully and be forced to write an unpublished decision. To analogize, settlement guarantees that the pool of cases appealed from the legalist trial court is stocked with fish. The higher stock must make the appellate court strictly better off. It ensures that the appellate court never wastes audit resources.

7.3 Appellate Precedent Doesn't Shift the Distribution of Cases

The model assumes that the precedent doesn't change the distribution of cases. It might be reasonable to think that cases "center" in the area of uncertain

law.[32] Maybe rather than always drawing cases from the unit interval, the cases truncate and are drawn uniformly on $x \in [\underline{x}, \widetilde{x}_t]$ – the area of uncertain law. We don't get more cases governed by clearly established law over time. Instead, all cases are just pushed into the area of uncertainty. Consider the legalist trial court in such a world. It makes its decision based on the average value of the precedent interval. It imposes liability on all cases above the average and no liability on cases below the average. Each period, then, the legalist trial court gets half the cases wrong from the appellate court's perspective (recall it wants liability imposed above $\underline{x}$). Indeed, in this thought experiment, the trial court gets the same "number" of cases wrong no matter the precedent bounds. So, in response, the appellate court always sets the audit probability at the same level. Precedent building, at first cut, doesn't seem to save on the appellate court's future audit costs.

We think that move is too quick. Our baseline model assumes a fixed loss to the appellate court from decisions by the trial court with which it disagrees. In terms of the size of the loss, it doesn't matter if the trial court finds the defendant not liable in a case next to $\underline{x}$ or a case far to the right of $\underline{x}$ – i.e., one where the activities costs are much higher. If the appellate court's loss turned on distance, the appellate court would benefit from using precedent to change the distribution of cases. Basically, even though the same number of cases arise, incorrect resolutions by the trial court matter less to the appellate court as it builds precedent. Suppose, say, that the cases are distributed in a small interval next to $\underline{x}$, given the precedent bounds. The appellate court doesn't care that much which way those cases are decided. Thus, it spends little checking the resolution in the trial court. On the other hand, suppose that the cases are distributed in a large interval next to $\underline{x}$. The legalist trial court will make the same number of no liability findings, but those findings will be in cases where the appellate court suffers a large loss. By tightening the interval of uncertain law, the appellate court reduces its exposure to no liability findings that it really disfavors.

7.4 Trial Court Fact Discretion

One key assumption in our model is that trial courts do not manipulate the facts. As a result, a successful appellate court audit always uncovers the location of

[32] For a model where the appellate court precedent alters the distribution of cases and this affects the appellate court's ability to learn, see (Parameswaran, 2014).

the actual case presented to the trial court, not some maniulated representation of that case. Some may question whether that account squares with actual practice. Appellate courts cannot set aside the trial courts findings of facts unless they are "clearly erroneous."[33] Knowing this standard, trial courts might manipulate facts if doing so serves their interests.[34] In a leading formal account of fact discretion, Gennaioli and Schleifer (2008) provide a model of how this freedom to characterize facts may affect the behavior of trial courts and appellate courts. In this subpart, we suggest how two of their most important assumptions could be applied to our model.

The Gennaioli and Schleifer model assumes that trial courts have discretion in the way they characterize the facts of the case. Quite sensibly, the model assumes that the cost to the trial court increases as the distance between its characterization of the facts and the "true" facts increases. When it comes to appellate courts, the model assumes that they must take the facts as the trial court reports them.

If the legalist trial court has some fact discretion, it is likely to use that discretion to minimize the amount of effort it has to expend to resolve cases. This court's inclination to use that discretion should turn on the relative cost of the analogizing cost versus the characterization cost. If one makes the extreme assumption that the cost to characterize cases as the trial court desires is infinitesimal, the trial court will characterize every case as identical to existing precedent. To return to the numerical example that began with precedent at .1 and .95, the trial court will characterize any case below .525 as .1 and any case above that threshold as .95. The appellate court would never audit because it will never get a chance to establish a new precedent. At the other extreme–when characterization costs are infinite–our model comes out exactly the same because the trial courts would never use their factual discretion. A more realistic scenario would be that analogizing costs and characterization costs are both moderate. Under these assumptions, the trial court will use its discretion to characterize facts until the marginal cost of doing so is equal to the marginal cost of analogizing.

The ability to shift the facts will also be attractive to the realist trial court insofar as it can characterize cases as consistent with the not liable precedent. To return to the example, the realist court will, where cost justified, depict the

[33] See Federal Rule of Civil Procedure 52(a)(6).

[34] The assumption in this subsection is that trial courts report a set of facts that the appellate court can observe only through a successful audit.

facts as consistent with the .1 precedent. The reason the realist court does so is that an audit of this case may not result in a reversal. Upon a successful audit, the appellate court might only observe the manipulated .1 facts and, consequently, affirm the liable disposition. The cut point at which the realist court will use factual discretion will be where the marginal cost of using that discretion equals marginal reduction of the expected loss that will come with a successful audit and a reversal.

We suspect that the ability for the trial court to exercise discretion may have some effect on how the law converges in our model. If one assumes that the appellate court either cannot uncover manipulated facts or that the appellate court must work harder to uncover these cases, the appellate court may have to make law on the basis of manipulated facts. To the degree, that trial court interests diverge from those of the appellate court, those manipulated facts may not be as good (from the perspective of the appellate court) for establishing new law relative to the "true" facts of the case. Relative to our model, this situation may mean that the appellate court needs either to work harder or audit more cases before the law eventually converges.

8 Conclusion

Appellate courts typically do two tasks. First, they review and correct dispositions in the trial courts. Second, they make new law to be applied by trial courts. The model we develop shows the interaction between the two tasks. Appellate scrutiny of trial court cases is necessary to find cases upon which new law can be made. Making new precedent changes the way trial courts resolve cases, enabling the appellate court to devote fewer resources to auditing those dispositions. Models in the literature typically consider either one task or the other. Such models have trouble explaining affirmances with opinions, the practice of dicta and how appellate scrutiny changes over time. By considering both tasks in a dynamic model, the model explains these common features of the judiciary as consistent with a dynamic optimization problem. We can also explain why trial courts focus so much on finding a case on point and why appellate court want to provide more of those cases.

The model leaves out important aspects of the legal system. We assume that all cases are appealed. There will be selection into appeal, which the appellate court audit strategy should account for and respond to. Second, the

trial courts are short-lived. They do not care how today's decision impacts the payoff to future trial courts. If they did, one might suppose they would flag cases likely to supply good precedent for analogical reasoning and thus reduce their future decision costs. Finally, the model treats the appellate court as a unitary actor. Issues of bargaining among judges over the contours of an opinion or the decision to publish were suppressed. We leave these questions for future work.

References

Abramowicz, M. and Stearns, M. (2004). Defining dicta. *Stan. L. Rev.*, 57:953.

Alexander, L. and Schauer, F. (1997). On extrajudicial constitutional interpretation. *Harv. L. Rev.*, 110:1359.

Baker, S. and Mezzetti, C. (2012). A theory of rational jurisprudence. *Journal of Political Economy*, 120(3):513–551.

Beim, D., Hirsch, A. V., and Kastellec, J. P. (2014). Whistleblowing and compliance in the judicial hierarchy. *American Journal of Political Science*.

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

Che, Y.-K. and Kartik, N. (2009). Opinions as incentives. *Journal of Political Economy*, 117(5):815–860.

Choi, S. J., Gulati, M., and Posner, E. A. (2012). What do federal district judges want? an analysis of publications, citations, and reversals. *Journal of Law, Economics, and Organization*, 28(3):518–549.

Cross, F. (2005). Appellate court adherence to precedent. *Journal of Empirical Legal Studies*, 2(2):369–405.

Daughety, A. F. and Reinganum, J. F. (2000). Appealing judgments. *The Rand Journal of Economics*, pages 502–525.

De Mesquita, E. B. and Stephenson, M. (2002). Informative precedent and intrajudicial communication. *American Political Science Review*, 96(04):755–766.

Dorf, M. C. (1994). Dicta and article iii. *University of Pennsylvania Law Review*, 142(6):1997–2069.

Frisch, D. (2003). Contractual choice of law and the prudential foundations of appellate review. *Vand. L. Rev.*, 56:57.

Gennaioli, N. and Shleifer, A. (2007). The evolution of common law. *Journal of Political Economy*, 115(1):43–68.

Gennaioli, N. and Shleifer, A. (2008). Judicial fact discretion. *J. Legal Stud.*, 37:1–577.

Holmes, O. W. (1870). Codes, and the arrangement of the law. *Am. L. Rev.*, 5:1.

James, F., Hazard, G., and Leubsdorf, J. (1992). *Civil procedure*. Little, Brown.

Jones, E. H. (1995). Back to the future for federal appeals courts: Rationing federal justice by recovering limited jurisdiction. *Tex. L. Rev.*, 73:1485–1821.

Katyal, N. K. (1998). Judges as advice givers. *Stan. L. Rev.*, 50:1709.

Kim, P. (2007). Lower court discretion. *New York University Law Review*, 82(2):383–442.

Klein, D. and Devins, N. (2013). Dicta, schmicta: Theory versus practice in lower court decision making. *William & Mary Law Review*, 54(6).

Kornhauser, L. A. (1992a). Modeling collegial courts i: Path-dependence. *International Review of Law and Economics*, 12(2):169–185.

Kornhauser, L. A. (1992b). Modeling collegial courts. ii. legal doctrine. *Journal of Law, Economics, & Organization*, 8(3):441–470.

Law, D. S. (2004). Strategic judicial lawmaking: Ideology, publication, and asylum law in the ninth circuit. *U. Cin. L. Rev.*, 73:817.

Lax, J. R. (2007). Constructing legal rules on appellate courts. *American Political Science Review*, 101(3):591.

Lax, J. R. (2012). Political constraints on legal doctrine: how hierarchy shapes the law. *The Journal of Politics*, 74(03):765–781.

Leval, P. N. (2006). Judging under the constitution: Dicta about dicta. *NYUL Rev.*, 81:1249.

Levy, M. K. (2013). Judicial attention as a scarce resource: A preliminary defense of how judges allocate time across cases in the federal courts of appeals. *Geo. Wash. L. Rev.*, 81:401–627.

Ljungqvist, L. and Sargent, T. J. (2012). *Recursive Macroeconomic Theory*. MIT Press.

McNollGast (1995). Politics and the courts: A positive theory of judicial doctrine and the rule of law. *Southern California Law Review*, 68:1631–1684.

Mead, R. A. (2001). "unpublished" opinions as the bulk of the iceberg: Publication patterns in the eighth and tenth circuits of the united states courts of appeals. *Law Library Journal*, 94(4):589–607.

O'hara, E. (1993). Social constraint or implicit collusion-toward a game theoretic analysis of stare decisis. *Seton Hall L. Rev.*, 24:736.

Parameswaran, G. (2014). Learning, opinion writing and the efficiency of the common law.

Posner, R. A. (2006). Reasoning by analogy. *Cornell L. Rev.*, 91:761–1383.

Rasmusen, E. B. (1994). Judicial legitimacy: An interpretation as a repeated game. *Journal of Law, Economics, and Organization*, 10:63–83.

Schanzenbach, M. M. and Tiller, E. H. (2008). Reviewing the sentencing guidelines: Judicial politics, empirical evidence, and reform. *The University of Chicago Law Review*, pages 715–760.

Schauer, F. (1995). Giving reasons. *Stan. L. Rev.*, 47:633.

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

Shavell, S. (2006). The appeals process and adjudicator incentives. *J. Legal Stud.*, 35:1–475.

Songer, D. R., Cameron, C. M., and Segal, J. A. (1995). An empirical test of the rational-actor theory of litigation. *Journal of Politics*, 57:1119–1129.

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

Sunstein, C. R. (1993). On analogical reasoning. *Harvard Law Review*, 106(3):741–791.

Appendix

Proof of Proposition 1

(A) Since $c'(1) = \infty$, it is obvious that $p_1^* \neq 1$, $p_2^* \neq 1$and $q_1^* \neq 1$; thus those constaints don't bind. The maximization problem is

$$\underset{p_1>0;\ p_2>0;\ q_1>0}{Max} \sum_{i=1}^{2} \beta_i \pi_i$$

where

$$\pi_1 = -(x^* - \underline{x})\lambda(1 - p_1) - x^* c(p_1) - (1 - x^*)c(q_1)$$
$$\pi_2 = -(1 - \underline{x})\lambda(1 - p_2) - c(p_2)$$

The Kuhn-Tucker conditions are

$$\frac{\partial L}{\partial q_1} = -\beta_1(1 - x^*)c'(q_1^*) \leq 0;\ q_1^* \geq 0\ ;\quad \frac{\partial L}{\partial q_1} q_1^* = 0$$
$$\frac{\partial L}{\partial p_1} = \lambda\beta_1(x^* - \underline{x}) - \beta_1 x^* c'(p_1^*) \leq 0;\ p_1^* \geq 0\ ;\quad \frac{\partial L}{\partial p_1}(\ p_1^*) = 0$$
$$\frac{\partial L}{\partial p_2} = \beta_2(1 - \underline{x})\lambda - \beta_2 c'(p_2^*) \leq 0;\quad p_2^* \geq 0\ ;\quad \frac{\partial L}{\partial p_2}(p_2^*) = 0$$

Suppose that q_1^* is in the interior (i.e., $q_i^* \in (0, 1)$). In that case, complementary slackness implies that $\frac{\partial L}{\partial q_1} = 0$ or

$$0 = -\beta_1(1 - x^*)c'(q_1^*)$$

a contradiction since $c'(q_1^*)$ is positive when $q_i^* \in (0, 1)$. Thus, $q_1^* = 0$. Suppose that $p_1 = p_2 = 0$ (so $c' = 0$). Rearranging results in

$$\lambda\beta_1(x^* - \underline{x}) \leq 0$$
$$\beta_2(1 - \underline{x})\lambda \leq 0$$

a contradiction. Thus, p_1^* and p_2^*. are at the interior. The second order condition holds since $c'' > 0$

Lemma 1. A value function $V(\widetilde{x}_t)$ exists. The value function is decreasing

and uniquely defined by

$$V(\widetilde{x}_t) = \max_{\substack{0 \le p_{1t} \le 1, 0 \le q_{1t} \le 1 \\ 0 \le p_{2t} \le 1,}} \sum_{i=1}^{2} \beta_i \pi_i$$

$$+\delta V(\widetilde{x}_t) \left([1 - \widetilde{x} + \underline{x}] + \beta_1 \{ (1 - p_{1t}) \int_{\underline{x}}^{x_1^*} dx + (1 - q_{1t}) \int_{x_1^*}^{\widetilde{x}} dx \} + \beta_2 (1 - p_{2t}) \int_{\underline{x}}^{\widetilde{x}} dx \} \right)$$

$$+ \delta \beta_1 \left(p_{1t} \int_{\underline{x}}^{x_1^*} V(x) dx + q_{1t} \int_{x_1^*}^{\widetilde{x}} V(x) dx \right) + \delta \beta_2 p_{2t} \int_{\underline{x}}^{\widetilde{x}} V(x) dx$$

Proof

To prove the lemma, we follow the approach in Baker & Mezzetti (2012), which applies the sufficiency conditions from Blackwell's Theorem (Blackwell 1965) to a similar problem (on this approach, see also Ljungqvist & Sargent (2012)).

Define S as the metric space of continuous functions mapping $x \in [\underline{x}, \overline{x}]$ into the real line. Consider the complete metric $d(v, w) = \sup_{x \in [\underline{x}, \overline{x}]} |v(x) - w(x)|$. Let T be an operator mapping continuous functions, v, into a new function, Tv. Specifically let

$$Tv = \max_{0 \le p_{1t} \le 1, 0 \le q_{1t} \le 1, 0 \le q_{1t} \le 1} \sum_{i=1}^{2} \beta_i \pi_i$$

$$+ \delta v(\widetilde{x}_t) \left([1 - \widetilde{x}_t + \underline{x}] + \beta_1 \{ (1 - p_{1t}) \int_{\underline{x}}^{x_1^*(\widetilde{x}_t)} dx + (1 - q_{1t}) \int_{x_1^*(\widetilde{x}_t)}^{\widetilde{x}_t} dx \} + \beta_2 (1 - p_{2t}) \int_{\underline{x}}^{\widetilde{x}_t} dx \} \right)$$

$$+ \beta_1 \left(p_{1t} \int_{\underline{x}}^{x_1^*(\widetilde{x}_t)} v(x) dx + q_{1t} \int_{x_1^*}^{\widetilde{x}_t(\widetilde{x}_t)} v(x) dx \right) + \beta_2 p_{2t} \int_{\underline{x}}^{\widetilde{x}_t} v(x) dx$$

Given v, Tv is the guess at the value function. We want to prove that Tv is a contraction mapping and, as a result, a unique and continuous value function exists. To do so, we must show monotonicity and discounting (Blackwell, 1965). Suppose that $v(x) \ge w(x)$ for all $x \in [\underline{x}, \overline{x}]$. It is immediate that $Tv > Tw$ and monotonicity holds (the integration of the $w(x)$ terms is always smaller than the integration of the $v(x)$ terms if $w(x) < v(x)$). M

To show discounting, take a constant c and a function $v + c$. We need to show that $T(v + c) = Tv + \gamma c$ where $\gamma \in [0, 1]$. We have

$$
\begin{aligned}
T(v+c) = & \max_{0 \le p_{1t} \le 1, 0 \le q_{1t} \le 1, 0 \le q_{1t} \le 1} \sum_{i=1}^{2} \beta_i \pi_i \\
& + \delta[v(\widetilde{x}_t) + c] \left([1 - \widetilde{x}_t + \underline{x}] + \beta_1 \{ (1 - p_{1t}) \int_{\underline{x}}^{x_1^*(\widetilde{x}_t)} dx + (1 - q_{1t}) \int_{x_1^*(\widetilde{x}_t)}^{\widetilde{x}_t} dx \} + \beta_2 (1 - p_{2t}) \int_{\underline{x}}^{\widetilde{x}_t} dx \} \right) \\
& + \delta \beta_1 \left(p_{1t} \int_{\underline{x}}^{x_1^*(\widetilde{x}_t)} [v(x) + c] dx + q_{1t} \int_{x_1^*}^{\widetilde{x}_t(\widetilde{x}_t)} [v(x) + c] dx \right) + \delta \beta_2 p_{2t} \int_{\underline{x}}^{\widetilde{x}_t} [v(x) + c] dx
\end{aligned}
$$

which equals

$$T(v + c) = Tv + \delta c$$

Discounting holds.

To show that the value function is negative, notice that if v is negative, then, Tv is negative. Finally, we need to show that Tv maps decreasing functions into decreasing functions. and, as a result, the value function increases in the amount of defined law. Suppose v is decreasing (i.e., $v' < 0$). Evaluated at the optimal values, we have

$$
\begin{aligned}
\frac{\partial T v(\widetilde{x})}{\partial \widetilde{x}} = \beta_1 \frac{\partial x_1^*}{\partial \widetilde{x}} & [-\lambda(1 - p_{1t}^*) - c(p_{1t}^*) + c(q_{1t}^*) + \delta(p_{1t}^* - q_{1t}^*)[v(x_1^*) - v(\widetilde{x})] \\
& + v' \{ 1 - \beta_1 p_{1t}^* [x_1^* - \underline{x}] - \beta_1 q_{1t}^* [\widetilde{x} - x_1^*] - \beta_2 p_{2t}^* [\widetilde{x} - \underline{x}] \} < 0
\end{aligned}
$$

We know that $\frac{\partial x_1^*}{\partial \widetilde{x}}, v < 0$, and $v' < 0$. The term in the curly brackets is always positive for all values of the parameters and choice variables (recall that $\beta_2 = 1 - \beta_1$)).The first term is negative if λ is sufficiently large. Note that the term is not necessarily negative for all values of λ. Because v is decreasing, the term $\delta(p_{1t}^* - q_{1t}^*)[v(x_1^*) - v(\widetilde{x})]$ can be positive (it will be if $p_{1t}^* > q_{1t}^*$. The condition amounts to saying that the appellate court prefers that trial court the case on the margin between liable and not liable, flip from not liable to liable. If it does, the appellate court gains some utility this period. At the same time, it loses with some probability the opportunity to learn about the marginal case (this happens if $p_{1t}^* > q_{1t}^*$). The conditions says that the lost opportunity to learn is outweighed by the gains in fewer loses in the current period. That will

always happen if λ is large enough.

Proof of Proposition 2

Since $c'(1) = \infty$, it is obvious that $p^*_{1t} \neq 1$, $p^*_{2t} \neq 1$and $q^*_{1t} \neq 1$; thus those constaints don't bind. The Kuhn Tucker conditions are

$$\frac{\partial Obj}{\partial q_{1t}} = \frac{\delta \int_{x_1^*}^{\widetilde{x}_t} (V(x) - V(\widetilde{x}_t)dx}{1 - x_1^*} - c'(q^*_{1t}) \leq 0; \quad q^*_{1t} \geq 0 : \quad \frac{\partial Obj}{\partial q_{1t}} q^*_{1t} = 0$$

$$\frac{\partial Obj}{\partial p_{1t}} = \frac{1}{x_1^*}\{\lambda(x_1^* - \underline{x}) + \delta \int_{\underline{x}}^{x_1^*} (V(x) - V(\widetilde{x}_t)dx - c'(p^*_{1t}) \leq 0; \quad p^*_{1t} \geq 0 : \quad \frac{\partial Obj}{\partial p_{1t}} p^*_{1t} = 0$$

$$\frac{\partial Obj}{\partial p_{2t}} = \lambda(1 - \underline{x}) + \delta \int_{\underline{x}}^{\widetilde{x}_t} (V(x) - V(\widetilde{x}_t)dx - c'(p^*_{2t}) \leq 0; \quad p^*_{2t} \geq 0 : \quad \frac{\partial Obj}{\partial p_{2t}} p^*_{2t} = 0$$

Suppose that $q^*_{1t} = 0$. This implies that

$$\frac{\delta \int_{x_1^*}^{\widetilde{x}_t} (V(x) - V(\widetilde{x}_t)dx}{1 - x_1^*} \leq 0$$

Since V is negative and decreasing, $\int_{x_1^*}^{\widetilde{x}_t} (V(x) - V(\widetilde{x}_t)dx$ is positive if $\widetilde{x}_t > x_1^*$ (i.,e., so long as the law hasn't converged).and we have a contradiction. So, $q^*_{1t} > 0$.

To prove part (b), at the interior solution, the first order conditions with respect to p_{1t} and q_{1t}. are

$$\frac{\partial Obj}{\partial p_{1t}} = \frac{1}{x_1^*}\{\lambda(x_1^* - \underline{x}) + \delta \int_{\underline{x}}^{x_1^*} (V(x) - V(\widetilde{x}_t)dx = c'(p^*_{1t})$$

$$\frac{\partial Obj}{\partial q_{1t}} = \frac{\delta \int_{x_1^*}^{\widetilde{x}_t} (V(x) - V(\widetilde{x}_t)dx}{1 - x_1^*} = c'(q^*_{1t})$$

We made the following assumption about the initial setting of the precedent

on liability: $\overline{x} = 1 - \underline{x}$. Any new liability precedent will always be set less than $\overline{x}$ (i.e., $\widetilde{x}_t < \overline{x}$). Because $x_1^* = \frac{\widetilde{x}_t + \underline{x}}{2}$, it is immediate that $x_1^* < 1 - x_1^*$ for all values of $\widetilde{x}_t$. Note that the denominator on the marginal benefit side (the LHS) is smaller in the first equation. Suppose that $p_{1t}^* < q_{1t}^*$. For that to be true, we need

$$\lambda(x_1^* - \underline{x}) + \delta \int_{\underline{x}}^{x_1^*} (V(x) - V(\widetilde{x}_t) dx < \delta \int_{x_1^*}^{\widetilde{x}_t} (V(x) - V(\widetilde{x}_t) dx$$

or

$$\begin{aligned}
\lambda(x_1^* - \underline{x}) &< \delta \int_{x_1^*}^{\widetilde{x}_t} (V(x) - V(\widetilde{x}_t) dx - \delta \int_{\underline{x}}^{x_1^*} (V(x) - V(\widetilde{x}_t) dx \\
&= \delta \int_{x_1^*}^{\widetilde{x}_t} (V(x) dx - \delta[\widetilde{x}_t - x_1^*] V(\widetilde{x}_t) dx - \delta \int_{\underline{x}}^{x_1^*} V(x) dx + \delta[x_1^* - \underline{x}] V(\widetilde{x}_t) dx \\
&= \delta \int_{x_1^*}^{\widetilde{x}_t} (V(x) dx - \delta \int_{\underline{x}}^{x_1^*} V(x) dx < 0
\end{aligned}$$

The first equality passes because $\widetilde{x}_t - x_1^* = x_1^* - \underline{x}$. The final expression is less than zero because (1) the integration is over the same range since x_1^* lies at the average of $\widetilde{x}_t$ and $\underline{x}$ and (2) $V(x)$ is decreasing in x. Since $\lambda(x_1^* - \underline{x}) > 0$, we have a contradiction. As a result, $p_{1t}^* < q_{1t}^*$.

Proof of Proposition 3

Totally differentiate the first order conditions with respect to $\widetilde{x}_t$ to obtain the comparative statics. Doing so yields

$$\frac{\partial p_{2t}^*}{\partial \widetilde{x}_t} = \frac{\delta \int_{\underline{x}}^{\widetilde{x}_t} dx V'}{-c''} > 0$$

$$\frac{\partial p_{1t}^*}{\partial \widetilde{x}_t} = \frac{\{c' - \lambda - \delta\{V(x_1^*) - V(\widetilde{x}_t)\} \frac{\partial x_1^*}{\partial \widetilde{x}_t} + \delta \int_{\underline{x}}^{x_1^*} dx V'}{-x_1^* c''}$$

The first inequality follows because $V' < 0$. In the second expression, the numerator is negative since $\lambda > c'$ by the assumption in the proposition; V is

decreasing and, as result, $V(x_1^*) > V(\widetilde{x}_t)$; $\frac{\partial x_1^*}{\partial \widetilde{x}_t} > 0$; and $V' < 0$.

Proof of Proposition 4

The appellate court loses the opportunity to make law if it stops auditing. To demonstrate that the appellate court makes law for all possible cases (i.e., the liability precedent eventually equals $\underline{x}$), we show that the audit probabilities will not equal zero unless $\widetilde{x}_t = \underline{x}$. Suppose that $p_{1t}^* = 0$. In that case, we must have

$$\frac{1}{x_1^*}\{\lambda(x_1^* - \underline{x}) + \delta \int_{\underline{x}}^{x_1^*} (V(x) - V(\widetilde{x}_t)dx \leq 0$$

a contradiction. Proofs for p_{2t}^* and q_{1t}^* are similar.

To prove part (B), note that, after convergence $x_1^* = \underline{x}$. Thus, the cutoff for the legalist trial court matches the cutoff for the appellate court. As a result, it doesn't make sense to spend any resources auditing the legalist trial court.

Proof of Proposition 5

At the interior solution, the first order conditions with respect to $\Delta(\cdot)$ for each $x \in [\underline{x}, x_1^*]$ and $x \in [x_1^*, \widetilde{x}_t]$ are respectively

$$\frac{\partial Obj}{\partial \Delta(x)} = \frac{\partial\{\alpha\Delta(x) + (1-\alpha)x\}}{\partial \Delta(x)} \delta V^{\Delta'}(\cdot)\{\beta_1 p_{1t}^* + \beta_2 p_{2t}^*\} = 0$$
$$\frac{\partial Obj}{\partial \Delta(x)} = \frac{\partial\{\alpha\Delta(x) + (1-\alpha)x\}}{\partial \Delta(x)} \delta V^{\Delta'}(\cdot)\{\beta_1 q_{1t}^* + \beta_2 p_{2t}^*\} = 0$$

By argument similiar to the one in proposition 4, all the audit probabilities will be positive so long as the law hasn't converged. By arguments similiar to the ones in lemma 1, if λ is sufficiently large, then, the value function decreases in x (i.e., $V^{\Delta'}(\cdot) < 0$). Thus, the only way for the first order conditions to hold is if

$$\frac{\partial\{\alpha\Delta(x) + (1-\alpha)x\}}{\partial \Delta(x)} = 0$$

Recall that $\alpha = \frac{\Delta(x) - x}{\widetilde{x}_t - \underline{x}}$. Thus, we can write the expression above as $\Delta^*(x)$

$$\frac{\partial\{\frac{\Delta(x)^2 - \Delta(x)\underline{x} + x - \Delta(x)x}{\widetilde{x}_t - \underline{x}}\}}{\partial \Delta(x)} = 2\Delta(x) - \underline{x} - x = 0$$

Solving, we see that $= \frac{\underline{x}+x}{2}$. Recall that x is the case drawn. It is the case value the appellate court can use to make law tied directly to the case facts. Since $\underline{x} < x$, it follows that $\Delta^*(x) < x$. The appellate court issues dicta; statements

not connected directly to the case facts.

To prove part (C), note that the legalist trial court suffers effort losses if the case lies in the interval $[\underline{x}, \widehat{x}_t]$. Otherwise there is a case plus dicta sufficiently close to be on point. By using dicta, the appellate court always reduces this interval.