Hard Cases Make Bad Law? A Theoretical Investigation

Sepehr Shahshahani
Fordham University School of Law, sshahshahani@fordham.edu

Hard Cases Make Bad Law?
A Theoretical Investigation

Sepehr Shahshahani*

Abstract

I use formal models to probe the aphorism that "hard cases make bad law." The analysis illuminates important features of the common law process, especially the influence of case characteristics on lawmaking and the role of strategic litigators. When a case raises concerns that are not reflected in doctrine, the court might distort the law to avoid a hardship. Distortion is more likely when the case is important or the facts are close to the border of legality. Litigators may exploit courts' attention to extra-doctrinal concerns by strategically selecting cases for litigation. Surprisingly, though, a strategic litigator improves lawmaking relative to random case selection—even when her preferences are far from the ideal rule—if her influence over case selection is modest. The effect is more nuanced when the strategic litigator has greater selection power. Finally, the analysis incorporates a judicial hierarchy with asymmetric information and factfinding discretion.

*I am especially grateful to the editor William Hubbard and two anonymous referees for comments that have helped greatly improve the manuscript, and to Lewis Kornhauser for his support and conversation. I also thank Pam Bookman, Rick Brooks, Chuck Cameron, Courtney Cox, John Ferejohn, Ezra Friedman, Martin Gelter, Lewis Kornhauser, Jae Lee, Henry Smith, Richard Squire, Ben Zipursky, and participants in the annual meeting of the American Political Science Association, the Emory Conference on Institutions and Lawmaking, the NYU Colloquium on Law, Economics, and Politics, the Northwestern Law and Economics Colloquium, and the annual conference of the Society for Institutional & Organizational Economics for helpful comments and discussions.

"Hard cases make bad law" is one of the most famous aphorisms in Anglo-American law. Courts and commentators are not always clear about what they mean by it, but its basic logic is as follows: Where strict application of a generally sound law would present a special hardship to someone, the court is tempted to distort the law to avoid the hardship. This paper presents a series of formal models of adjudication that make this logic precise and probe it, revealing nuances that would not be appreciated by informal analysis alone. There is some intrinsic interest in unpacking this oft-quoted maxim. More importantly, the analysis uses the aphorism as a vehicle to illuminate larger questions about judicial lawmaking. A fundamental feature of the common law process is that general law is made in the context of particular cases, such that courts' lawmaking and dispute-settling functions are inextricable. The insights of this paper come from appreciating this linkage and what it implies—namely, that a few cases' particular characteristics can exert great influence on generally applicable law. Given the motivation, I will focus on characteristics that are usefully classified under the rubric of hardness.

Three senses of "hard" are considered: (1) *special hardship*, meaning salient facts that cannot be explicitly reflected in relevant doctrine; (2) *importance*, meaning not the importance of the law at issue but the importance of the dispute; (3) *difficulty* or *closeness*, meaning how close the facts of the case are to running afoul of the law. I show that when a case does not pose a special hardship, importance and difficulty do not make a difference to lawmaking quality. But when a case poses a special hardship, important cases are more likely than unimportant cases, and difficult cases are more likely than easy cases, to make bad law. However, conditional on making bad law, difficult cases make less-bad law than easy cases.

As mentioned, these results flow from the critical linkage of lawmaking and dispute-settling in the common law process. I also investigate another critical feature of judicial lawmaking—that courts make law by resolving cases brought to them *by others*. This highlights the role of "impact litigators" (also called "cause lawyers"), meaning lawyers whose primary goal is not that a particular client should win but that the law should take a particular shape. Impact litigators can move the law closer to their liking by strategically selecting cases for litigation with an eye to how particular case characteristics affect general laws. I show that this process is more conducive to good lawmaking the closer the litigator's preferred rule is to the hypothetical ideal rule. Surprisingly, though, impact litigation improves lawmaking relative to random case selection—even when the litigator's preferred rule is maximally far from the ideal rule—as long as the litigator has modest influence over case selection. The intuition is that in filtering out cases that would make bad law from her own perspective, the impact litigator also tends to filter out cases that would make bad law from the perspective of the ideal rule; in expectation, her case selection circumvents the tendency of courts to let hard cases make bad law. This logic holds as long as the impact litigator's power over case selection is modest—that is, as long as she is not free to hold off until she finds a case that will settle the law very close to her ideal rule.

Impact litigators' impact on lawmaking quality is more nuanced when their influence over case selection is greater. The analysis identifies three distinct regimes, depending on the proximity of the litigator's ideal rule to the ideal rule. For litigators with strong prosocial preferences, society is always better off with impact litigators than without, and increasing selection power always improves lawmaking. For litigators with moderately prosocial preferences, society is still always better off with impact litigators than without; however, increasing selection power is welfare-enhancing up to a point but becomes welfare-reducing after that point. For litigators with extreme preferences, increasing selection power is welfare-enhancing up to a point but becomes welfare-reducing after that point; moreover, though modest selection power is better than having no impact litigator, having no impact litigator is better than an extremely powerful litigator.

I also explore whether hard cases make bad law in a judicial hierarchy where trial courts have some discretion in factfinding. Trial courts are better positioned than appellate courts to observe case facts, and appellate courts defer substantially to their findings of fact. Informational asymmetry, trial courts' strategic factfinding, and appellate courts' strategically responsive rulemaking complicate the single-court analysis. Unlike the single-court context, in a judicial hierarchy some cases that do not pose a special hardship may make bad law. Moreover, the effect of difficulty is nonmonotonic: The cases that are most likely to make bad law are intermediately difficult—not the easiest cases but not the hardest cases either. Finally, when bad law is made, it is bad not only in the sense of diverging from the appellate court's ideal rule but in the stronger sense of being Pareto-dominated for both courts.

This work relates closely to two strands of literature. The first is the "case space" approach pioneered by Kornhauser (1992) and now common in judicial politics (see Lax (2011) for a review). In this framework, the court decides a case, which consists of a set of facts, by announcing a rule that partitions the fact space; cases falling on one side of the rule get one disposition (plaintiff wins) and those on the other side get another disposition (plaintiff loses). The great virtue of Kornhauser's framework is its recognition that lawmaking (deciding the governing law) and dispute settling (deciding the case's disposition) are different but inextricably linked. My analysis exploits this key feature by showing how various characteristics of a particular dispute shape the law that is made in the context of that dispute but will apply more broadly. Other papers have also fruitfully exploited this potentiality of case space. For example, Carrubba and Clark (2012) integrate both rule and disposition components into a court's payoff function, and Shahshahani (2021) shows how trial courts' fact discretion sharpens appellate courts' rule-disposition tradeoff.

In the case space literature, Lax (2012) is close to the present work in investigating lawmaking in an environment of imperfect doctrine by modeling a second dimension of facts that is not fully incorporated into law. But the two works focus on different aspects of lawmaking. Lax (2012) allows for improving the visibility of the second dimension by costly investment, and more generally for two-dimensional doctrine, which enables him to investigate the choice between a rule and a standard. In my model, by contrast, doctrinal imperfection is so severe that the second dimension simply cannot be reflected in doctrine (at a non-prohibitive cost), so the rules-versus-standards tradeoff is not considered. The central tradeoff in my basic model, in line with its purpose of investigating how specific case characteristics influence general lawmaking, is between the right disposition in the present case and the right rule for future cases. By contrast, particular case characteristics are not important to lawmaking in Lax (2012); "salience" in that model means the salience of the issue area, not of a particular case, which impacts how much the court invests in lawmaking but does not create any sort of rule-disposition tradeoff. Beyond the benchmark model, my analysis of impact litigators has no analogue in Lax (2012), and my analysis of judicial hierarchy is different in fully modeling lower courts as strategic actors.

The second strand of related literature is the body of work on the evolution of common law. Since early in the inception of law and economics, scholars have claimed that the common law gropes toward efficiency (e.g., Posner (1973), Rubin (1977), Priest (1977)). Though some of the early works could be criticized for exceedingly strong assumptions (see Kornhauser (1980)), they contained insights that were incorporated into more sophisticated models of litigation that generated somewhat optimistic, though more modest and refined, results (e.g., Cooter, Kornhauser and Lane (1979), Cooter and Kornhauser (1980), Gennaioli and Shleifer (2007), Baker and Mezzetti (2012)). The present work is not necessarily more pessimistic about the trajectory of common law, but it focuses on the conditions that make the law go awry. Three distinguishing features enhance this focus. First, where this literature emphasizes the malleability of the law, I emphasize its stickiness. Here the law is really law in the sense that future judges cannot change it at will. Of course, both approaches contain stylized assumptions, and one or the other may be more appropriate depending on the context. (Gennaioli and Shleifer (2007) take a middle ground where future judges cannot change one dimension of the law but can introduce a second dimension, at which point the law as a whole becomes fixed.) Second, the present paper is attentive to how the characteristics of a particular case influence general law. In the common law efficiency literature, case characteristics are not modeled; cases are generic objects, distinguished, if at all, only by whether they cross the line into illegality. But in the present model, in line with longstanding lawyerly intuitions, the court is liable to be influenced by the particular attributes of the parties and the case before it—even attributes that are not technically relevant under the law. That is why there is room for bad lawmaking even though I incorporate the strong assumption of a benevolent and knowledgeable court, an assumption that is lacking in some of the more refined common law models. Third, building on the second point, the model introduces an analysis of how strategic litigation impacts the quality of lawmaking through selection of cases to litigate. Baker and Biglaiser (2014) also model impact litigation, but they focus on whether impact litigators will seek incremental or dramatic change and do not consider how particular case characteristics might affect the law.

Section 1 reviews the origins and usage of the aphorism. Section 2 constructs a series of formal models to clarify, enrich, contest, and extend the qualitative insights. Section 3 discusses implications and extensions with case studies from diverse areas of law, pulling in substantive and procedural doctrine. Among other things, I show how the model complements the traditional justification for equity. Apart from avoiding the hardship to particular parties that would result from strict application of law, equity preserves good law by giving courts a way to avoid the hardship without having to distort the law. Section 4 concludes.

1 Origins and Usage

The aphorism was apparently first used in English cases in the early 19th Century (see Heuston (1978)). In *Hodgens v. Hodgens* (1837) 4 CI Fin. 323, a husband whose wealthy wife had left him petitioned the court for maintenance of their children out of her property, claiming his own resources were insufficient. The law at the time did not impose any duty on the mother while the father was alive. Nevertheless, a court in Dublin ruled in favor of the father so the children would not become destitute. The House of Lords reversed, Lord Wynford remarking as follows: "We have heard that hard cases make bad law. This is an extremely hard case, but it would indeed be making bad law ... if your Lordships affirmed this order" (*id.* at 378). Lord Wynford expressed the hope that the wife would still feel bound by "the law of God and nature ... suitably to maintain those children"; but, as far as courts are concerned, "we have to decide this case according to the law" (*id.* at 377-78).

In *Winterbottom v. Wright* (1842) 10 M. & W. 109, a coach driver who was injured in an accident sued the person who had furnished the coach, claiming the accident was caused by latent defects in the coach. The court dismissed the suit because any duty to keep the coach safe was owed to the person who contracted for the coach to be furnished, not to the driver. Baron Rolff wrote, "it is, no doubt, a hardship upon the plaintiff to be without a remedy, but, by that consideration we ought not to be influenced. Hard cases, it has been frequently observed, are apt to introduce bad law" (*id.* at 116).

A modern case nicely illustrates the same idea. Cindy Lee Garcia sued YouTube to take down the trailer for a movie in which she had briefly appeared, claiming she owned the copyright in her performance. The suit would be a sure loser under ordinary circumstances given Garcia's minimal contribution to the film. It's well settled that authorship of a "joint work" under the Copyright Act resides only in a creative "master mind" with "artistic control" over the work, which "limit[s] authorship [in movies] to someone at the top of the screen credits, sometimes the producer, sometimes the director, possibly the star, or the screenwriter" (*Aalmuhammed v. Lee*, 202 F.3d 1227, 1233 (9th Cir. 2000)). The rationale is that a more inclusive standard of authorship would impose immense transaction costs, especially in movies, which involve huge production teams (*id.*). But these were no ordinary circumstances. The film at issue was *Innocence of the Muslims*, which depicted the Muslim prophet Muhammad as a murderer and pedophile and caused worldwide protests. Garcia had no idea what she was getting into when she answered a casting call for an action flick set in the Arabian desert, and her lines were later dubbed over to insult Muhammad. She received numerous death threats after the trailer was posted on YouTube. In view of Garcia's grave predicament, a panel of the Ninth Circuit twisted copyright doctrine to rule in her favor (*Garcia v. Google, Inc.*, 766 F.3d 929 (9th Cir. 2014)).

Hodgens, *Winterbottom*, and *Garcia* capture the essence of the aphorism: When a case presents a special hardship, the court is tempted to bend a generally sound law to avoid the hardship, resulting in a law that, though perhaps fine for the case at hand, is unsound as a general rule. That is how most commentators understand the maxim (e.g., Garner (2011), 403; Heuston (1978), 31; Radin (1938), 40-42).[1]

But that does not express all there is to the qualitative insight. Consider also *Northern Securities Co. v. United States*, 193 U.S. 197 (1904), involving an antitrust challenge by President Theodore Roosevelt's administration to the merger of competing railroad companies that resulted in the formation of the Northern Securities holding company, the world's largest company at the time. The Supreme Court invalidated the combination, holding that mergers between directly competing firms are per se illegal. Justice Holmes wrote in dissent (*id.* at 400-401):

> Great cases, like hard cases, make bad law. For great cases are called great, not by reason of their real importance in shaping the law of the future, but because of some accident of immediate overwhelming interest which appeals to the feelings

[1]Schauer (2006) runs farther with the idea, arguing that because cases (not just cases posing a special hardship) are unrepresentative of the range of problems that the law would be called upon to resolve, case-by-case lawmaking makes bad law.

> and distorts the judgment. These immediate interests exercise a kind of hydraulic pressure which makes what previously was clear seem doubtful, and before which even well settled principles of law will bend.

This is the most-cited articulation of the maxim. What it adds to the *Hodgens-Winterbottom-Garcia* intuition is the idea of a case's "great"ness or importance. A case posing a special hardship always tempts judges to distort the law to avoid the hardship, but the temptation is easier to resist when the hardship is local. Hence the opinions in *Hodgens* and *Winterbottom*, which recognize the distortionary temptation only to rebuff it. In *Garcia*, too, the Ninth Circuit took the case en banc and reversed the panel decision, apparently able to resist the distortionary pressure given that the hardship was imposed on only one person (*Garcia v. Google, Inc.*, 786 F.3d 733 (9th Cir. 2015) (en banc)). When the hardship relates to a pressing public concern, however, the temptation is harder to resist. Hence, as Holmes saw it, the majority's succumbing to antimonopoly sentiment and veering from the sound path of law in *Northern Securities*. (Of course, this is not to endorse Holmes's position in *Northern Securities*, much less the rules of *Hodgens* and *Winterbottom*; the point is rather to get a feel for judicial usage of the aphorism.)

These cases, taken together, capture the meaning of "hard cases make bad law" as generally used.[2] Judicial usage, then, suggests two senses of a "hard" case: a case that poses a special hardship and a case whose outcome is particularly important.

But there is a third sense of "hard" in legal discourse, meaning "difficult" (or "close"), the opposite of "easy." A hard case in this sense is a case that is not readily resolvable by reference to precedent or other authorities. As Ronald Dworkin put it in a famous article of the same title, "hard cases" are those in which "no settled rule dictates a decision either way" (Dworkin (1975), 1060). The question of how judges should go about deciding hard

[2]This was confirmed by canvassing all references to the aphorism in Supreme Court opinions. E.g., *United States v. Clark*, 96 U.S. 37, 49 (1877) (Harlan, J., dissenting); *Allen v. Morgan Cnty.*, 103 U.S. 515 (1880); *F.C.C. v. WOKO, Inc.*, 329 U.S. 223, 229 (1946); *Dennis v. United States*, 341 U.S. 494, 528 (1951) (Frankfurter, J., concurring); *New York Times Co. v. United States*, 403 U.S. 713, 753 (1971) (Harlan, J., dissenting); *id.* at 759 (Blackmun, J., dissenting); *Nixon v. Admin. Gen. Serv.*, 433 U.S. 425, 505 (1977) (Burger, C.J., dissenting); *Green v. Georgia*, 442 U.S. 95, 98 (1979) (Rehnquist, J., dissenting); *Skinner v. Ry. Labor Exec. Ass'n*, 489 U.S. 602, 654-55 (1989) (Marshall, J., dissenting); *Hilton v. S.C. Pub. Ry. Comm'n*, 502 U.S. 197, 207 (1991) (O'Connor, J., dissenting); *Board of Cnty. Comm'rs v. Umbehr*, 518 U.S. 668, 710 (1996) (Scalia, J., dissenting); *Caperton v. A.T. Massey Coal Co.*, 556 U.S. 868 (2009) (Roberts, C.J., dissenting); *Kansas v. Carr*, 136 S. Ct. 633, 651 (2016) (Sotomayor, J., dissenting).

cases is central to the field of jurisprudence (e.g., Dworkin (1975); Posner (2002); Shapiro (2007)) and also surfaces in judges' writing (e.g., Sutton (2010), who helpfully uses the term "close" cases). To my knowledge, the aphorism has never been used to mean "difficult cases make bad law."[3] Nevertheless, one may wonder how the difficulty of a case influences the quality of lawmaking, and how difficulty interacts with the other two senses of hardness. These questions will be addressed in my formal analysis.

2 Models

Having surveyed the usage of judges and commentators, we have rich enough intuition to build on. Sections 2.1-2.3 take on board the qualitative intuitions and sharpen them with the aid of some formalism, showing the effects of special hardship, importance, and difficulty. Subsequent sections go beyond the choice-theoretic context assumed by the aphorism, using game theory to analyze impact litigators (§§ 2.4-2.5) and judicial hierarchy (§ 2.6).

2.1 Building Blocks

We now analytically conceptualize every word of the saying.

Hard. All three senses of hardness are considered. Special hardship is conceptualized as a latent dimension of case facts that, for whatever reason (e.g., administrability or evidentiary considerations), cannot be explicitly reflected in legal doctrine. Importance is conceptualized as how much the court cares about the case's disposition. (Note well that this refers to the intrinsic dispositional importance of the case, in line with Holmes's notion of "immediate overwhelming interest," not its doctrinal importance.) And difficulty is conceptualized as the closeness of the case to the court's ideal cutpoint rule.

Cases. In accordance with the case space approach, a case is modeled as a bundle of

[3] However, Justice Stevens was fond of "easy cases make bad law." E.g., *Burnham v. Superior Court*, 495 U.S. 604, 640 (1990); *Ankenbrandt v. Richards*, 504 U.S. 689, 718 (1992); *Hudson v. United States*, 522 U.S. 93, 106 (1997). In using this variation, Justice Stevens seemed to think he is inverting the traditional expression, which is not right because the opposite of "hard" in the traditional usage is not "easy." Other Justices have also used the variation and attempted to articulate a rationale for it—e.g., *O'Bannon v. Town Court Nursing Center*, 447 U.S. 773, 804 (1980) (Blackmun, J., concurring); *Heckler v. Chaney*, 470 U.S. 821, 840 (1985) (Marshall, J., concurring)—but pursuing this line of thought would take us too far afield.

facts. Or, more precisely, of facts that are or should be legally relevant. Going back to the *Hodgens* case, the *law* (or *rule*) specifies the wife and husband's respective shares of financial child-maintenance duties. (Specifically, the *Hodgens* court's rule was that the wife's share of duties is 0, but one can imagine a rule imposing any share.) The case would then consist of what share of maintenance was actually borne by the wife in the controversy before the court, and the court's rule would generate a *disposition* of the case, meaning a determination of whether or not the wife is in compliance with her legal duties. For higher-dimensional rules (e.g., if the allocation of maintenance duties also depended on the spouses' wealth), the case would be modeled as a higher-dimensional bundle of facts. More precisely, a case x is a point in fact space $X \subset \mathbb{R}^n$, and a rule r is a hyperplane dividing the fact space into two half spaces, each corresponding to a disposition $d \in \{0, 1\}$.

Make. The idea of a case making law presupposes that the rule of the case will apply beyond that case. In *Hodgens*, for example, the court cannot just say the wife wins (or loses), but must specify an allocation of maintenance duties—a rule—that makes the wife win (or lose). Moreover, this rule applies not just to the Hodgenses but to future parties as well.

Bad. Badness captures the extent to which the dispositions produced by the law over the expected run of future cases overlap with the dispositions that would have been produced by the court's ideal rule. For example, in one-dimensional fact space, where rules take the form of cutpoints, the badness of a rule is captured by how close it is to the ideal rule.[4] I do not offer a normative theory of what constitutes good or bad law; rather, taking the normative benchmark of good law as given, and assuming that the court is interested in attaining it, I show when and to what extent hard cases will cause the court to deviate from good law.

Law. As discussed above, a "law" means a rule that divides the fact space into half spaces corresponding to two dispositions (win or lose).

Utility function. Given the qualitative discussion, we must make clear that the court cares both about the disposition of the case at hand and about the rule made by the case

[4]I will assume throughout that case facts are uniformly distributed over the fact space, as is common in the literature (e.g., Lax (2003), Gennaioli and Shleifer (2007); Beim, Hirsch and Kastellec (2014)). A similar exercise could be carried out for an arbitrary distribution (with appropriate assumptions about its support and density), but the uniform distribution is simple and allows for a clean focus on the effects of hardness.

(i.e., the disposition of future cases). In one-dimensional fact space, this is captured by

$$U = -|r - H| + e\mathbb{1}\{d = d_H\} \tag{1}$$

where r is the rule of the case, H is the court's ideal rule, and e is the dispositional payoff, which accrues if and only if the disposition of the case (d) conforms to the court's ideal disposition (d_H). Ideal disposition means the disposition demanded by the court's ideal rule:

$$d_H = \begin{cases} 1 & \text{if } \; x < H \\ 0 & \text{if } \; x \geq H \end{cases}. \tag{2}$$

So the first term in Equation (1) is the court's rule utility and the second term is disposition utility, with e capturing the case's importance (i.e., the ratio of disposition utility to rule utility). Note that the court's payoff function reflects a certain myopia. A judge with a long horizon—meaning one who is concerned about all future cases and potential cases—would not care about the disposition of the particular case at hand, because the importance of one case pales in comparison with the great run of future cases. However, in line with the aphorism (and longstanding lawyerly intuitions), the judge in the model cannot abstract away from the circumstances of this case and focus only on the law that is best for all cases. That is why the judicial payoff function contains not only the rule payoff ($-|r - H|$) but also a disposition payoff (e) for getting the particular case right.

To summarize: In the single-court models that follow, the court decides a case by choosing a rule, which generates its payoff as per equation (1).

2.2 Single Court with Perfectly Inclusive Doctrine

First consider a rule that incorporates all relevant facts (a "perfectly inclusive" rule). In other words, there are no "special hardships" of which the law cannot take account. Then the court's unique optimal action would be to set the rule at its ideal point ($r = H$ in one dimension, the choice of the ideal separating hyperplane in higher dimensions). This rule is the unique maximizer of the rule component of the court's utility function, and it also guarantees the correct disposition. The choice is uniquely optimal irrespective of case importance or case difficulty. In particular, case importance does not matter because, given perfectly inclusive doctrine, the ideal rule always generates the correct disposition. This is all obvious; the purpose is simply to establish a benchmark for later analyses:

Remark 1. *In the single-court context, if doctrine is perfectly inclusive then cases never make bad law.*

2.3 Single Court with Under-Inclusive Doctrine

Next consider a rule that cannot reflect all relevant factual dimensions. For example, the actor's plight in *Garcia* was important but could not be incorporated into copyright authorship doctrine. There are many reasons why courts may not be able to incorporate every conceivably relevant factor into law. For example, a higher lawmaking body (e.g., the legislature) may have cabined the factors that courts can consider. Or it might be that incorporating additional factors would make the law too complicated to provide meaningful notice to those who would be expected to comply with it, or would create too many loopholes for clever actors looking for avenues of noncompliance, or would be too costly for administrability or evidentiary reasons. Whatever the reason might be, it is common that doctrine cannot take cognizance of all conceivably relevant facts, and that is when the maxim has bite.

In particular, suppose that perfectly inclusive doctrine would take account of facts in two dimensions (x_1 and x_2), but it is feasible for doctrine to only consider facts in one dimension (x_1). To develop intuition, suppose the fact space is the unit square and the court's ideal perfectly inclusive rule (the "first-best rule") is $x_1 = x_2$, yielding ideal dispositions

$$d_H = \begin{cases} 1 & \text{if } x_1 < x_2 \\ 0 & \text{if } x_1 \geq x_2 \end{cases}. \tag{3}$$

But, given under-inclusiveness, the court must choose a rule of the form $r = x_1$, yielding

$$d = \begin{cases} 1 & \text{if } x_1 < r \\ 0 & \text{if } x_1 > r \end{cases}. \tag{4}$$

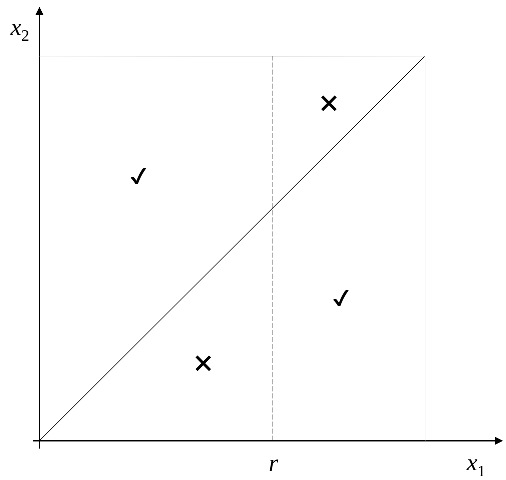


Figure 1: The ideal rule ($x_1 = x_2$) takes cognizance of facts in two dimensions, but the feasible rule r can take cognizance of facts on only one dimension (x_1).

(To avoid epsilon problems, assume the court can choose either disposition when $x_1 = r$.) Clearly, then, doctrine is unavoidably imperfect: For any rule choice, some cases will not get the disposition dictated by the first-best rule, as in the regions marked $\times$ in Figure 1.

Given that doctrine is under-inclusive and the first-best rule unfeasible, the court's rule utility is maximized by the "second-best rule"—i.e., the rule that would decide the largest possible mass of cases in accordance with the first-best rule. (In Figure 1, pick r to maximize the regions marked ✓. Formally, maximize $\Pr(d = d_H)$.) Simple calculus verifies that the second-best rule is $r = 1/2$. The question of whether hard cases make bad law can now be understood as whether difficult or important cases are more likely (than easy or unimportant cases) to cause the court to deviate from its second-best rule, resulting in unnecessary distortion away from the first-best.

First consider case importance. A case comes before the court. The court must decide whether to choose the second-best rule or deviate from it to get the right disposition, hence making bad law. The court would be willing to sacrifice the second-best rule if and only if $e > |x_1 - 1/2|$.[5] That is, the court would be willing to make bad law if x_1 falls in the interval $(1/2 - e, 1/2 + e)$, but not outside it. (Even inside this interval, a case may not pose a rule-disposition tradeoff,[6] in which case the court need not make bad law, though it would be willing to do so if the tradeoff were posed.) So an increase in case importance increases

[5] This formulation uses $-|x_1 - 1/2|$ as a reduced form for the court's rule utility (expected future disposition utility). The expression can be microfounded by calculating expected future dispositional utility from two-dimensional cases as a function of different one-dimensional rules, which shows that rule utility is indeed single-peaked and symmetric around a maximum at $r = 1/2$ (though the microfounded form of the loss function is quadratic rather than linear).

[6] I.e., if $x \in \{(x_1, x_2) \,|\, x_1 \in (1/2 - e, 1/2) \text{ and } x_1 < x_2\} \cup \{(x_1, x_2) \,|\, x_1 \in (1/2, 1/2 + e) \text{ and } x_1 > x_2\}$.

the probability of making bad law, as shown in Figure 2.

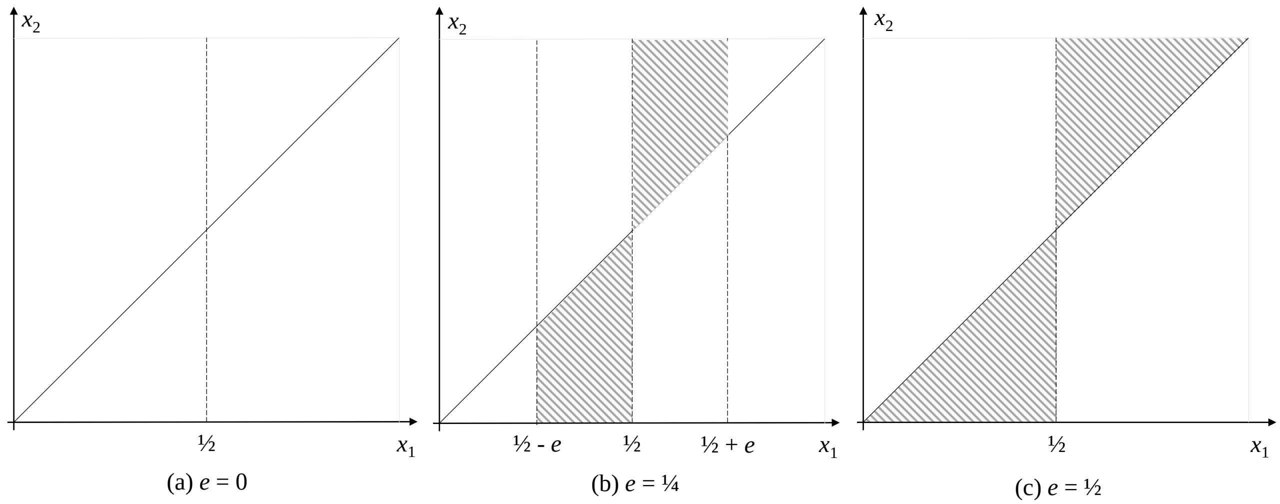


Figure 2: Cases making bad law (shaded regions) for three levels of case importance: (a) unimportant ($e = 0$), (b) moderately important ($e = 1/4$), (c) very important ($e = 1/2$).

Next consider case difficulty. This can be conceptualized as the distance between the legally-articulable dimension of case facts and the second-best rule ($|x_1 - 1/2|$), which captures the idea that "close" cases could go the other way if the case facts were just a little bit different. (Keep in mind that shorter distance means greater difficulty.) It is clear from Figure 2 (panels (b)-(c)) that difficult cases are more likely to make bad law. When the first dimension of case facts is closer to the second-best cutpoint, (1) the probability of conflict between the first-best and second-best dispositions is higher, so the case is more likely to pose a rule-disposition tradeoff, and (2) when such a tradeoff is posed, the court can achieve its preferred disposition by a smaller deviation from the second-best rule; therefore, the court is more likely to make bad law. (Formally, $\Pr(r \neq 1/2 \,|\, X_1 = x_1)$ is decreasing in $|x_1 - 1/2|$. That is, for $x_1 < 1/2$, $\frac{\partial}{\partial x_1} \Pr(X_2 < X_1 \,|\, X_1 = x_1) > 0$, and for $x_1 > 1/2$, $\frac{\partial}{\partial x_1} \Pr(X_2 > X_1 \,|\, X_1 = x_1) < 0$.)

The analysis so far has clarified three points. First, a case posing a special hardship can make bad law, capturing the intuition in *Hodgens* and other cases. Second, when a case poses a special hardship, important cases are more likely than unimportant cases to make bad law, capturing Holmes's intuition in *Northern Securities*. Third, when a case poses a special hardship, difficult cases are more likely than easy cases to make bad law, a relationship that has not been considered before.

Formal analysis allows us to say still more. The qualitative intuition is that hard cases make bad law, but how bad? Consider again, for any level of case importance, the region of case facts that would make bad law. Within that region, as Figure 2 makes clear (middle and right panels), more difficult cases actually make *less*-bad law. That is so because, when case facts along the legally-articulable dimension (x_1) are close to the second-best cutpoint, the court can flip the disposition of the case by only a small deviation from the second-best rule. By contrast, when case facts are far from the second-best cutpoint (easy cases), the level of rule distortion necessary to flip the disposition is high. Conditional on making bad law, more difficult cases make less-bad law. (There is no analogous effect for case importance; conditional on making bad law, the degree of badness does not change in case importance.)

The insights from the single-court context can be summarized as follows.

Proposition 1. *When one court makes law by deciding a case,*

1. *Cases not posing a special hardship never make bad law.*
2. *Among cases that pose a special hardship, more important cases are more likely to make bad law.*
3. *Among cases that pose a special hardship, more difficult cases are more likely to make bad law.*
4. *Conditional on making bad law, more difficult cases make less-bad law.*

2.4 Impact Litigators

The results of the previous section flow from the common law's intertwining of courts' dispute-settling and lawmaking functions. Another essential feature of judicial lawmaking is that courts make law by deciding cases brought to them *by others*. This highlights the potential for litigators' strategic case selection to alter the course of law. I now introduce a game to explore this topic. My object is to show how the law may be shaped by lawyers and activists who take interest in lawsuits not out of concern for a particular client but with an eye to developing general law. The importance of such impact litigators (aka cause lawyers), both on the left and on the right, is widely acknowledged.[7] Impact litigators have a vision of what they want the law to be, and survey the field to select a case that is apt to realize their vision (e.g., Hartocollis (2017)). In making this selection, they take advantage of judges' proclivity to let the particular facts of a case influence the making of general law, so the analysis dovetails with Section 2.3.

The players are a court (C) and an impact litigator (L). The court makes law by deciding a case. Its decision is guided, as before, by both rule utility and disposition utility. As in Section 2.3, the case has two factual dimensions, x_1 and x_2, only the first of which can be reflected in doctrine, as well as an importance dimension e. The impact litigator has a role in determining which case comes before the court as the vehicle for general lawmaking. She understands that the vehicle matters—i.e., that the resulting rule (r) might be different depending on case characteristics (x_1, x_2, e)—and wants to select a case that would produce a law close to her ideal rule. Of course, the litigator's ideal rule (r_L) may be different from the court's (H), so the litigator is not curating cases with an eye to developing "good" law (as the court understands that to be). Moreover, unlike the court, the litigator does not care how a particular case comes out; all she cares about is the law that the case would make. (This is not to say that the impact litigator would advocate against the interests of her client in a case, which would pose problems of legal ethics. Rather, as the formal model makes clear, the impact litigator's strategic calculation goes to the decision of what case to take.)

Sequence of play is as follows:

1. Nature draws a case (x_1, x_2, e) according to F_{X_1}, F_{X_2}, F_E. L decides whether or not to bring the case. If L brings the case then C decides the case by choosing a rule r. If L does not bring the case then the game proceeds to the second stage.

2. Nature draws a case (x_1, x_2, e) according to F_{X_1}, F_{X_2}, F_E and C decides the case by choosing r.

[7] Prominent examples of left-leaning impact litigation include the NAACP's efforts to end state-sanctioned racial segregation in the South during the mid-Twentieth Century and the ACLU's continued efforts to shape the laws pertaining to immigration, race, and sexual orientation (see, e.g., Tushnet (1987), Epp (1998), Mack (2012)). On the right, prominent examples include the use of litigation in tandem with other strategies to advance the deregulatory thrust of corporate and antitrust law in the late-Twentieth Century and litigation to roll back affirmative action (see, e.g., Teles (2008), Weinrib (2016), Hartocollis (2017)).

Payoffs are as follows:

$$U_C = -|r - H| + e\mathbb{1}\{d = d_H\} \tag{5}$$

$$U_L = -|r - r_L| \tag{6}$$

Equation (5) is the familiar judicial payoff function from Section 2.3, incorporating both rule and disposition utility. Equation (6) captures the long horizon of impact litigators, who care only about the rule and not about any particular disposition.[8]

In the model, impact litigators have a role in selecting cases, but their role is limited. They can "take a pass" on one bad draw of a case, but they cannot hold off indefinitely until the ideal case comes along. The idea is that the environment is rife with cases and people want their cases heard; impact litigators can influence which case will be the one that makes law, but they do not have a monopoly over litigation. If they pass up their opportunity for strategic selection, a randomly selected case will determine the law.

The outcome of interest is the quality of lawmaking (or "welfare"), which is conceptualized by reference to the *rule* utility of the court, disregarding its disposition utility. Formally, welfare is defined as expected equilibrium deviation from good law: $W = E\{-|r - H|\}$. As in section 2.3, this equation takes the court's ideal rule as the measure of "good" law. One may think of the welfare benchmark as the payoff function of a hypothetical judge who shares the court's view of what law is best but does not share the court's myopia or preoccupation with the particular case at hand; it's the payoff function of a philosopher king of the world with a long horizon (which is how Justice Holmes in *Northern Securities* and the Lords and Barons in *Hodgens* and *Winterbottom* appeared to think of themselves).

As before, I focus on a court with the first-best cutpoint $x_1 = x_2$, yielding the second-best rule $H = 1/2$. I assume as before that case facts are distributed uniformly over the unit square and case importance is distributed uniformly over $[0, 1/2]$ (recall that if $e = 1/2$ then the court is always willing if necessary to sacrifice the rule to the disposition). The model in

[8] Of course, a purely law-motivated litigator is an ideal type. One can conceive of purely client-motivated and purely law-motivated lawyers more generally as endpoints of a spectrum. In this more general model, the litigator's payoff function takes the form $U_L = -|r - r_L| + e_L\mathbb{1}\{d = d_L\}$, where d_L denotes L's ideal disposition (defined the same way as d_H in equation (2)) and e_L represents the weight on short-term client-centered motivation relative to long-term impact motivation.

this section can be solved for any given cumulative distribution function (with appropriate differentiability and continuity assumptions), but the uniform distribution has the advantage of simplicity and a clean focus on the effects of strategic case selection.

First benchmark: no impact litigator. If there is no impact litigator then the game has only the second stage, which is equivalent to the model in Section 2.3. The court's optimal strategy is to pick its ideal rule ($r = 1/2$) whenever there is no conflict between rule and disposition utility or the case is insufficiently important; and to distort the rule to the minimum extent necessary to achieve its preferred disposition ($r = x_1$) whenever there is a rule-disposition conflict *and* the case is sufficiently important. That is,

$$r = \begin{cases} x_1 & \text{if } \ x_1 > 1/2 \ \text{ and } \ x_2 > x_1 \ \text{ and } \ e > x_1 - 1/2 \\ x_1 & \text{if } \ x_1 < 1/2 \ \text{ and } \ x_2 < x_1 \ \text{ and } \ e > 1/2 - x_1 \\ 1/2 & \text{otherwise} \end{cases} \tag{7}$$

The resulting rule is shown in Figure 3. As Figure 3b shows, most cases make good law but some cases make bad law. In expectation there is some distortion of the ideal rule, and the expected magnitude of this distortion can be calculated to be 1/48.[9]

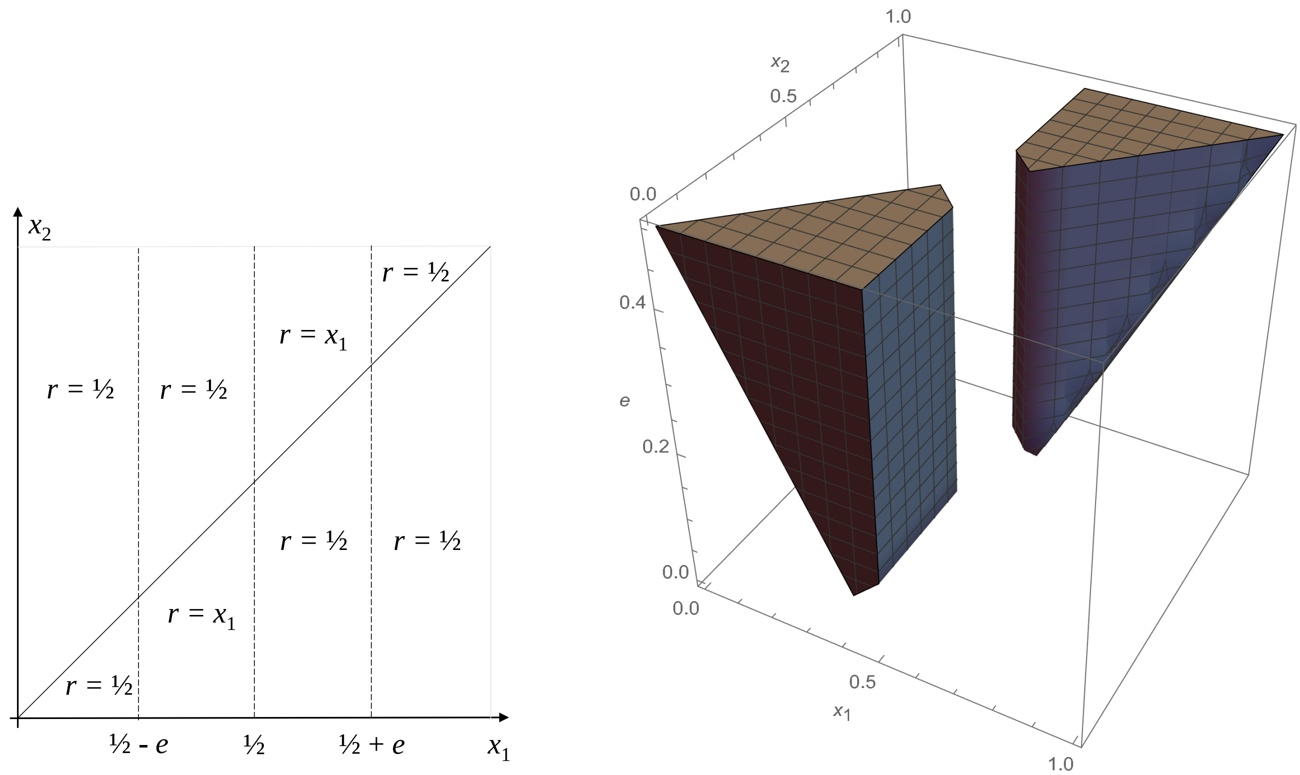
Second benchmark: ideal impact litigator. When the impact litigator shares the court's view of the ideal rule (i.e., $r_L = 1/2$), the litigator's payoff function is the same as the court's, except that the litigator is not myopic about case disposition. That is, the litigator has the same payoff function as the hypothetical long-horizon judge whose view is the benchmark of welfare.

Solving backwards: When a case is brought, the court decides it according to the decision rule of equation (7). So the impact litigator's expected payoff from moving to the second stage is $-1/48$ (the same as expected welfare in the stage-two game). Accordingly, in the first stage, the litigator brings the case drawn by Nature if and only if her expected payoff

[9] Formally,

$$W = \int_0^{0.5} \int_0^{x_1} \int_{0.5-x_1}^{0.5} x_1 - 0.5 dF_E(e) dF_{X_2}(x_2) dF_{X_1}(x_1) + \int_{0.5}^{1} \int_{x_1}^{1} \int_{x_1-0.5}^{0.5} 0.5 - x_1 dF_E(e) dF_{X_2}(x_2) dF_{X_1}(x_1)$$

which, given our distributional assumptions, equals $-1/48$.



(a) equilibrium rule for a fixed e

(b) “bad law” regions in parameter space

Figure 3: Panel (a) shows the equilibrium rule as a function of x_1 and x_2 for a fixed value of e. The filled part in panel (b) shows the regions in the entire parameter space where bad law ($r \neq 1/2$) would result.

from the case exceeds $-1/48$. This does *not* mean that the litigator brings a case if and only if $|x_1 - 1/2| < 1/48$. Indeed the litigator brings a case if that inequality is satisfied (i.e., if the first dimension of facts is sufficiently close to her ideal point, regardless of the second factual dimension and case importance); but the litigator *also* brings a case if the case does not pose a rule-disposition tradeoff or if it is insufficiently important to the court, because in both those scenarios the ideal rule ($r = 1/2$) would result. The impact litigator's equilibrium strategy is specified by the following decision rule:

$$L \text{ does } \textit{not} \text{ bring case iff} \begin{cases} x_1 < 23/48 \;\text{ and }\; x_2 < x_1 \;\text{ and }\; e > 1/2 - x_1 \\ \textit{or} \\ x_1 > 25/48 \;\text{ and }\; x_2 > x_1 \;\text{ and }\; e > x_1 - 1/2 \end{cases} \tag{8}$$

Under this decision rule, most configurations of case facts and case importance would lead the litigator to bring the case, and the probability that the game ends in the first stage is high. This is because most cases would make the rule $r = 1/2$, which is the litigator's ideal rule and, of course, preferable to proceeding to the second stage; even among cases that would not produce good law, some have first-dimension facts so close to the ideal rule that the rule distortion is smaller than the expected rule distortion in stage two (i.e., $|x_1 - 1/2| < 1/48$).

The upshot is that lawmaking quality is higher in the game with an "ideal" impact litigator than in the model without an impact litigator. It's easy to see why. The litigator can take a pass on one round of litigation, and she uses this power to not bring some hard cases that would make bad law. Because the litigator's conception of "bad" law is the same as the welfare benchmark, her strategic case selection improves welfare.

General case: any impact litigator. Consider, without loss of generality, an impact litigator whose ideal rule is to the right of the court's ($r_L > 1/2$). Again the game is solved backwards, and the court's equilibrium decision rule is given by equation (7). This time, however, the litigator's welfare is not the same as general welfare. If the game proceeds to the second stage, the litigator's payoff is given by

$$U_L^2 = \begin{cases} x_1 - r_L & \text{if } \ x_1 < 1/2 \ \text{ and } \ x_2 < x_1 \ \text{ and } \ e > 1/2 - x_1 \\ x_1 - r_L & \text{if } \ x_1 > 1/2 \ \text{ and } \ x_2 > x_1 \ \text{ and } \ e > x_1 - 1/2 \ \text{ and } \ r_L \geq x_1 \\ r_L - x_1 & \text{if } \ x_1 > 1/2 \ \text{ and } \ x_2 > x_1 \ \text{ and } \ e > x_1 - 1/2 \ \text{ and } \ r_L < x_1 \\ 1/2 - r_L & \text{otherwise} \end{cases} \tag{9}$$

Based on this, multiple integration shows that the litigator's expected second-stage utility is

$$EU_L^2 = -\frac{r_L^4}{3} + \frac{4r_L^3}{3} - 2r_L^2 + \frac{r_L}{3} + \frac{1}{6} \tag{10}$$

(Note that this expression equals $-1/48$ when $r_L = 1/2$. Note further that the impact litigator's expected second-stage payoff is decreasing in the distance between her ideal rule and the court's ($\partial EU_L^2/\partial r_L < 0$).)

In the first stage, the litigator brings a case if and only if her payoff from doing so is greater than the expected second-stage payoff in equation (10). Determining when this will occur is somewhat involved (see the Appendix), but the logic is similar to the second benchmark

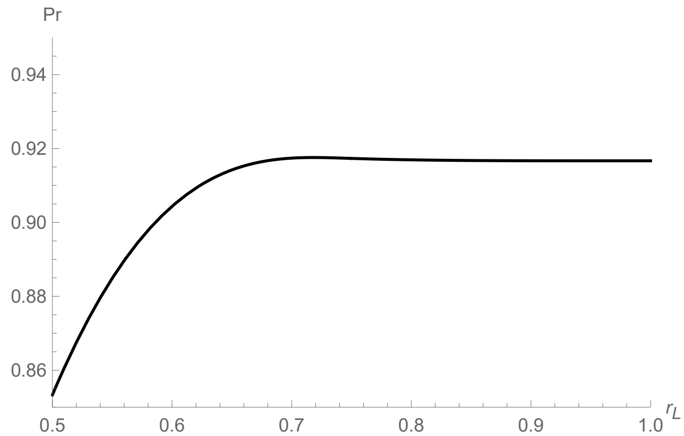


Figure 4: The probability that the impact litigator will bring a case in the first stage as a function of the litigator's ideal rule r_L.

model analyzed above. The cases which the litigator might bring fall into two categories: (1) cases the court would decide by making good law ($r = 1/2$), (2) cases the court would decide by making bad law ($r \neq 1/2$, which implies $r = x_1$). It turns out that *all* litigators would bring *any* case in the first category. (That is to say, $1/2 - r_L \geq EU_L^2 \; \forall r_L \in [1/2, 1]$.) As for the second category, there is an interval of first-dimension facts around the litigator's ideal rule for which she litigator is willing to bring the case. The bounds of this interval, which I denote $[\underline{x}_1, \overline{x}_1]$, move with the litigator's ideal rule. The lower bound is always below 1/2 (equaling 1/2 when $r_L = 1$) and the upper bound increases with r_L until it reaches 1 for sufficiently large r_L (see equation (A4) and associated discussion in the Appendix).

Figure 4 shows the probability that a case will be brought in the first stage as a function of the impact litigator's ideal rule. The function is nonmonotonic. The probability initially increases as the litigator's ideal rule diverges from the court's ideal rule, but once the divergence has become sufficiently large (roughly, for $r_L > 0.72$), the probability declines in ideal-rule divergence. (The probability is strictly decreasing in this region, though the rate of decline is so low that the graph appears like a horizontal line.)

Ultimately, the analysis in this section reveals an impact litigator's impact on welfare, defined as expected equilibrium deviation from good law. The solid curve in Figure 5 depicts the relationship between the litigator's ideal rule and welfare. Not surprisingly, welfare declines as the litigator's ideal rule diverges from the court's ideal rule. (The function is strictly decreasing, though for high values of r_L the rate of decline is so low that the curve appears horizontal.) Interestingly, though, expected welfare is always higher with an impact litigator (solid curve) than without (dashed line), even when her ideal rule is maximally far

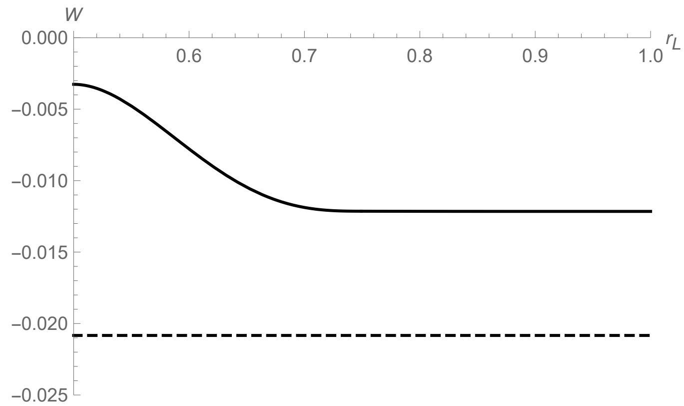


Figure 5: Expected equilibrium welfare as a function of the impact litigator's ideal rule r_L (solid curve) and expected equilibrium welfare without an impact litigator (dashed line).

from the court's. That is because, like the hypothetical philosopher-king judge and unlike the actual judge in the game, the impact litigator has a long time horizon. She cares only about the law, not about the disposition of the case that makes the law, and in serving this long-term interest she often selects cases that do not pose a rule-disposition tradeoff for the court. In other words, the impact litigator promotes the promulgation of good laws by strategically selecting cases that avoid the dynamic of "hard cases make bad law." The benefits of such strategic case selection outweigh the "drift" costs imposed by the impact litigator's desire to locate the rule as close as possible to her own preferred rule.

Proposition 2 summarizes the insights from the game with an impact litigator.

Proposition 2. *When the impact litigator has one pass at case selection,*

1. *The quality of lawmaking is higher with an impact litigator than without, regardless of the distance between the litigator's ideal rule and the socially ideal rule.*

2. *Lawmaking improves as the distance between the impact litigator's ideal rule and the socially ideal rule decreases.*

3. *The distance between the impact litigator's ideal rule and the socially ideal rule has a nonmonotonic effect on the probability that the litigator will select a case for litigation at the first stage.*

2.5 Impact Litigators with More Selection Power

The previous section considered an impact litigator who has one "pass" at selecting a case before Nature selects a case for the court to decide and make law. But imagine a model where the litigator could look at more than one draw from the case space before having to surrender case selection to Nature. In such a model the number of "passes" afforded to the litigator captures her selection power. A natural question then is how the quality of lawmaking changes as the litigator's selection power increases: Do emerging laws become monotonically worse? Are we better off without an impact litigator?

This section answers these questions. I make the previous section's two-period model more complex by considering an n-period model where the impact litigator can choose sequentially from $n-1$ cases before Nature pushes a case to the court. On the other hand, I simplify the model by considering one dimension instead of three. Cases are chosen randomly from the line, and the rule of the case is assumed to be its location. So the assumption is that Nature selects only hard cases (hence the court's locating the rule at the case facts), and the question is *which* hard case will be chosen to make the rule. In other words, *all* cases (except for a measure-zero point) make bad law, and the question is the *degree* of badness. The simplification both makes the calculations tractable and brings a sharper focus on the effect of selection power. The insights from the one-dimensional model travel unambiguously to the three-dimensional model for extremely high or low levels of selection power (n), but the effects of selection power in the intermediate range are more cleanly graspable in the one-dimensional model.

Sequence of play is as follows:

(1) Nature draws a case x according to F_X. L decides whether or not to bring the case. If L brings the case then C decides it by choosing $r = x$ and the game ends. If L does not bring the case then this case disappears and the game proceeds to the next stage.

⋮

$(n-1)$ Nature draws a case x according to F_X. L decides whether or not to bring the case. If L brings the case then C decides it by choosing $r = x$ and the game ends. If L does not bring the case then this case disappears and the game proceeds to the next stage.

(n) Nature draws a case x according to F_X. C decides the case by choosing $r = x$ and the game ends.

As before, $X \sim U[0,1]$ and the litigator's payoff and the welfare benchmark are

$$UL_n = E\{-|r - r_L|\} \tag{11}$$

$$W_n = E\{-|r - 1/2|\} \tag{12}$$

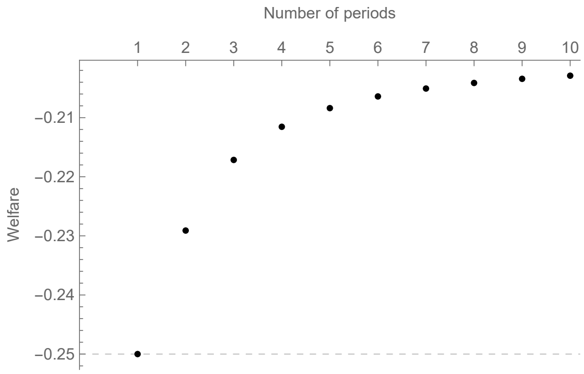
where the subscript n denotes the expected payoff of an n-period game.

In analyzing this game it is useful to define, for an n-period game, the equilibrium expected distance from the litigator's ideal rule and from the welfare benchmark, which I call D_n and DW_n respectively. (So $D_n = -UL_n$ and $DW_n = -W_n$.) We are interested in how these quantities change with n.

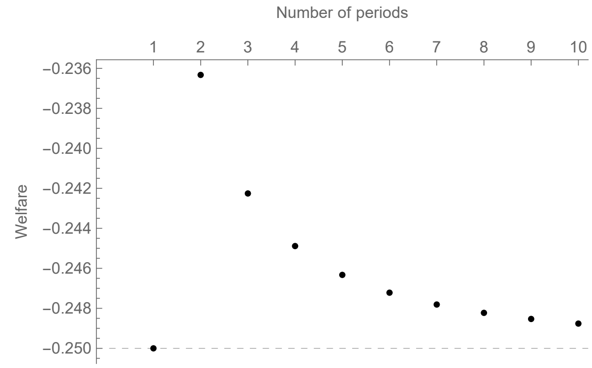
With respect to the litigator's payoff, the intuition is straightforward: As selection power increases, the expected distance between the equilibrium rule and the litigator's ideal rule decreases. What is more, the expected distance becomes arbitrarily small as the litigator's number of passes becomes arbitrarily large. (Formally, I show that the sequence (D_n) is decreasing and converges to 0.)

With respect to the welfare impact, it turns out, as in Section 2.4, that a little bit of selection power is always preferable to no selection power: Welfare improves when we move from a setting with no impact litigator to one with an impact litigator who has one pass at case selection, even for the most extreme litigator. (That is, $W_2 \geq W_1 \, \forall r_L$.) The intuition is that a litigator's interest in a rule close to her own ideal rule also works against the establishment of rules that are far from the socially optimal median rule; even litigators with extreme preferences to one side of the optimal rule enhance social welfare by vetoing cases that would make a rule close to the other extreme.

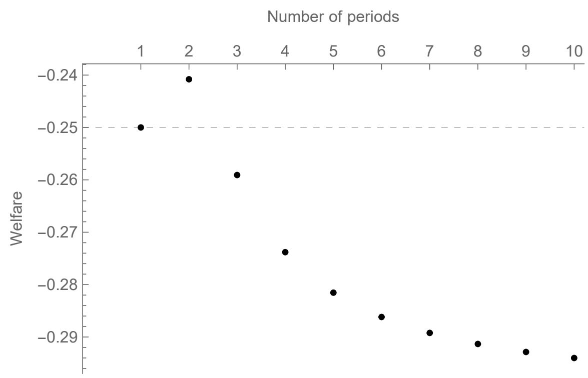
But does an increase in selection power continue to enhance welfare at higher levels of selection power? And is the impact litigator's presence always preferable to her absence, even if she has great selection power? The answers to these questions depend in nuanced ways on the distance between the impact litigator's ideal rule and the socially optimal rule. Three different regions of litigator preferences yield three different sets of answers (Figure 6). For litigators with preferences very close to the socially ideal rule (Figure 6a), having an impact litigator with *any* number of passes is preferable to not having one at all. What is



(a) $r_L = 0.7$



(b) $r_L = 0.75$



(c) $r_L = 0.8$

Figure 6: Expected equilibrium welfare as a function of impact litigator's selection power (number of periods in the game), for three different litigators: (a) litigator with preferences close to the ideal rule ($r_L = 0.7$), (b) litigator with preferences moderately far from the ideal rule ($r_L = 0.75$), (c) litigator with preferences far from the ideal rule ($r_L = 0.8$).

more, welfare always improves as the litigator's selection power increases. For litigators with preferences that are intermediately close to the ideal rule (Figure 6b), it is no longer true that increasing selection power is always beneficial; rather, increasing selection power is welfare-improving up to a point but becomes welfare-reducing after that point. However, we are still worse off without an impact litigator than with an impact litigator with *any* number of passes (even after increases in selection power have begun to erode welfare compared to lower levels of selection power). For litigators with preferences far from the ideal rule (Figure 6c), just as with intermediate impact litigators, increases in selection power are welfare-enhancing up to a point and welfare-reducing afterwards. But, unlike in the previous case, for high levels of selection power we are better off without an impact litigator.

The intuition behind these results is as follows. The impact litigator always uses her selection power to veto cases that would make a bad rule from her perspective, bringing the expected equilibrium rule closer to her own ideal rule. When the litigator's ideal rule is very close to the socially ideal rule (Figure 6a), bringing the expected equilibrium rule closer to her ideal rule also brings it closer to the socially ideal rule. But when the litigator's ideal rule is farther away from the socially ideal rule, the effects of increased selection power are no longer unambiguously good. In the beginning, increases in selection power are beneficial because the litigator will use her enhanced selection power to weed out extreme cases (i.e., cases that would make a rule far from both her ideal rule and the socially ideal rule). But, as selection power grows, the litigator can afford to be more discriminating; in addition to extreme cases, she begins to weed out some cases that are bad for her agenda but not so bad for social welfare. That is why, for impact litigators with intermediate or extreme preferences (Figure 6b - 6c), expanding selection power is welfare-improving only up to a point. Nevertheless, if the impact litigator's preferences are not very far from the socially ideal rule (Figure 6b), society is always better off with such an impact litigator, even if arbitrarily powerful, than without. By contrast, if the litigator's preferences are very far from the social ideal (6c), giving her ever greater selection powers can make society worse off, even compared to a setting with no impact litigator. Society is better off with no case curation at all than with an extremely powerful impact litigator with extreme preferences.

Proposition 3 summarizes the insights from the multiperiod impact-litigator game.

Proposition 3. *In the n-period game with an impact litigator,*

1. *The impact litigator's expected payoff is increasing in her selection power. Moreover, as the impact litigator's selection power becomes arbitrarily large, the expected rule gets arbitrarily close to her ideal rule. Formally, $D_{n+1} < D_n \,\forall\, n$ and $(D_n) \to 0$.*

2. *Society is better off with any impact litigator who has one pass than without an impact litigator. Formally, $W_2 > W_1 \,\forall\, r_L \in (0,1)$ and $W_2 = W_1$ when $r_L \in \{0,1\}$.*

3. *When the impact litigator's ideal rule is very close to the socially ideal rule, society is better off with an impact litigator than without, and an increase in the litigator's number of passes always improves social welfare. Formally, $\exists\, d$ such that, for all $r_L \in B_d(1/2)$, $W_n > W_1 \,\forall\, n > 1 \quad$ and $\quad W_{n+1} > W_n \,\forall\, n$.*

4. *When the impact litigator's ideal rule is intermediately close to the socially ideal rule, society is better off with an impact litigator than without; however, an increase in the*

litigator's number of passes improves social welfare up to a point and reduces social welfare after that point. Formally, for all r_L such that $|r_L - 1/2| \in [d, 1/4]$,

   (a) $W_n > W_1 \,\forall\, n > 1$ *and*

   (b) $\exists\, n'$ *such that* $W_{n+1} > W_n \,\forall n \leq n'$ *and* $W_{n+1} < W_n \,\forall n > n'$
   (but in the special case of $|r_L - 1/2| = d$ *we have* $W_{n'} < W_{n'+1} = W_{n'+2} = ...$*).*

5. *When the impact litigator's ideal rule is far away from the socially ideal rule, society is better off with an impact litigator than without iff the impact litigator has a sufficiently small number of passes; moreover, an increase in the litigator's number of passes improves social welfare up to a point and reduces social welfare after that point. Formally, for all r_L such that $|r_L - 1/2| > 1/4$,*

   (a) $\exists\, n'$ *such that* $W_{n+1} > W_n \,\forall\, n \leq n'$ *and* $W_{n+1} < W_n \,\forall n > n'$ *and*

   (b) $\exists\, n''$ *such that* $W_n < W_1 \,\forall\, n \geq n''$*.*

2.6 Judicial Hierarchy

The aphorism implicitly presupposes a single court making law by deciding a case, but in fact judiciaries are hierarchically structured. In the Online Appendix I continue the theme of investigating the impact of case characteristics on lawmaking, this time in the context of a judicial hierarchy. The focus is on an intermediate appellate court's review of a trial court's decision, where the trial court's legal determinations are reviewed *de novo* but its factual determinations are reviewed under the deferential "clear error" standard, the idea being that trial courts are better positioned to observe facts. The model, based on Shahshahani (2021), shows how informational asymmetry and trial courts' factfinding discretion moderate the influence of case characteristics on lawmaking. Briefly stated, the main results are as follows. First, in contrast with the single-court context (Remark 1), bad law may be made even if doctrine is perfectly inclusive and there is no special hardship. That is because the trial court's ability to slant the facts supplies an additional source of potential rule-disposition conflict for the lawmaking appellate court, which may lead to rule distortion. Second, the effect of case difficulty (closeness) on lawmaking is nonmonotonic. The cases that may make bad law are intermediately difficult—not the easiest cases (those with facts very far from the appellate court's ideal rule) nor the most difficult ones (those with facts very close to the appellate court's ideal rule). The intuition is that the appellate court's rule-disposition tradeoff is heavily weighted on side or the other in both easy and difficult cases, so the appellate court either willingly accepts the trial court's factfinding or shows such hostility that it deters the trial court from certain kinds of factfinding, in both cases obviating any rule distortion; but the rule-disposition tradeoff is finely balanced in the region of intermediate difficulty, so there is a mixed-strategy equilibrium where rule distortion occurs with positive probability in this region. Third, when bad law is made, the law is bad not only in the sense of deviating from the appellate court's ideal rule but in the stronger sense of being Pareto-dominated for both appellate and trial courts. In other words, there are rules that both courts would prefer to the equilibrium rule. That is because, for the appellate court's threat of rule distortion to deter the trial court from fudging the facts, the threatened rule must be *farther* from the trial court's ideal rule than is the appellate court's ideal rule (otherwise, rule distortion would benefit rather than punish the trial court). A full analysis, including a literature review and formal statements and proofs, may be found in the Online Appendix.

3 Implications, Applications, Extensions

3.1 Field Distortion and Justiciability

The Introduction and Section 1 gave examples of cases that fit the "hard cases" adage and my "missing dimension" model of it. Beyond individual cases, there are entire *fields* that seem susceptible to this dynamic. These are fields in which doctrinal issues often arise in cases involving parties who are sympathetic or unsympathetic for reasons unrelated to the central purpose of the doctrine. There is perhaps no clearer illustration of such field distortion than the Supreme Court's Fourth Amendment jurisprudence. The Fourth Amendment protects people against unreasonable searches and seizures. This protection, held the Warren Court, embodies an "exclusionary rule" that makes evidence uncovered by violating the Fourth Amendment inadmissible in criminal prosecutions (e.g., *Mapp v. Ohio*, 367 U.S. 643 (1961)). Fourth Amendment doctrine for the past half century or more has developed almost exclusively in the context of whether inculpatory evidence is to be excluded from a criminal trial—with the consequence that citizens' privacy rights have been eroded because of judges' natural reluctance to hold that "the criminal is to go free because the constable has blundered" (*People v. Defore*, 242 N.Y. 13, 21 (1926) (Cardozo, J.)).

For example, the "third party doctrine" holds that a person does not have a "reasonable expectation of privacy" in information voluntarily conveyed to third parties, so the government's accessing such information does not constitute a "search" or "seizure" triggering the Fourth Amendment's application. The rule was originally developed in cases where a criminal defendant had confided incriminating information to an undercover informant (e.g., *On Lee v. United States*, 343 U.S. 747 (1952); *Lopez v. United States*, 373 U.S. 427 (1963)). But it would later operate to erase Fourth Amendment protection for entire categories of personal information that citizens routinely and without much practical choice disclose to private parties (e.g., *United States v. Miller*, 425 U.S. 435 (1976) (a person has no Fourth Amendment interest in his bank records); *Smith v. Maryland*, 442 U.S. 735 (1979) (installation of pen register on phone was not a "search" under Fourth Amendment)).

The trend of courts shrinking Fourth Amendment rights for fear of letting criminals off the hook illustrates the principle that the nature of the case that makes the law matters to what law is made. It follows that the distribution of justiciable cases in a field affects the shape of the field. So procedural doctrines like standing that control what types of cases the courts may hear affect what kind of law courts will make—and the effect is greater than is commonly supposed. It is easy to see that if a particular kind of plaintiff does not have standing to bring suit, the kinds of issues faced by that kind of plaintiff will go unredressed. But the effect is broader than that. It's not just that the law will not now address the issues posed by a certain kind of case; it's that this kind of case will not be in the mix of cases that the court uses to make law in this field—so the court may in the end make law that addresses this kind of case too, but it will make that law using other kinds of cases. The resulting law may be dramatically different if justiciable and nonjusticiable cases are systematically different along the three axes of hardness.

For example, the Supreme Court held in *Clapper v. Amnesty International*, 568 U.S. 398

(2013), that a group of lawyers, human rights advocates, and media organizations did not have standing to challenge the 2008 amendments to the Foreign Intelligence Surveillance Act. Even though the plaintiffs had alleged that they regularly communicated with foreign persons whom the FISA amendments, but not prior law, authorized the government to surveil (e.g., families of Guantánamo Bay detainees or targets of the CIA's extraordinary renditions), the Court held that the prospect of surveillance was too "speculative" to confer Article III standing. Critics have pointed out that if there is no standing in this case, surveillance could go on for a long time without its constitutionality being examined by an Article III court. What is more (and that is the insight here), Fourth Amendment law, including law that bears on the constitutionality of FISA amendments, will continue to be made in *other* cases—cases involving accused criminals or terrorists seeking to suppress evidence of wrongdoing, where the immediate consequences of finding a Fourth Amendment violation will be felt far differently than when the target of the alleged violation is a human rights advocate seeking to uncover torture or abuse, so resulting protections will be less robust.

3.2 Impact Litigators' Impact

One of the counterintuitive results of the formal analysis is that impact litigators with modest selection power improve lawmaking, even when their preferences are far from the ideal rule. Bearing in mind the inherent limitations of any effort to provide an illustration for such a result,[10] a plausible example may be found in *Javins v. First National Realty Corp.*, 428 F.2d 1071 (D.C. Cir. 1970). The opinion is a milestone in landlord-tenant law (e.g., Rabin (1984), Chused (2004)). It held that residential leases are subject to an implied warranty of habitability and that the tenant's obligation to pay rent depends on the landlord's obligation to keep the premises habitable. Much remains debatable about the reforms ushered in by *Javins* and similar cases: The extension of contractually inalienable legal protections to

[10]It is difficult to provide an illustration of this result, for two reasons. First, impact litigators are often active in ideologically charged areas, so an uncontroversial example is hard to find; lefwingers tend to like letftwing impact litigators and rightwingers the rightwing kind. Second, impact litigation becomes well-known when it is most successful—which is to say, when the impact litigator's influence over case selection is large, not when it is modest. More generally, not all counterintuitive results of formal analysis lend themselves well to nice real-world examples, especially in partial-equilibrium analysis. One of the virtues of rigorous theory is to lay bare forces that are not empirically observable, or at least not readily so (because their effects are masked by other forces or for some other reason).

residential leases has been praised for protecting vulnerable tenants unable to individually negotiate lease terms or criticized for hurting those it aims to help by resulting in higher rents and a restricted supply of housing (Merrill and Smith (2001)). At a minimum, though, *Javins* was right to recognize that a residential lease is not purely a conveyance of an interest in land but also a contractual transaction subject to some contract law principles. So, just as it would be inefficient to hold that a buyer whose cow has not been delivered cannot raise that fact as a defense if sued for nonpayment and must sue separately to get his benefit of the bargain (*Nichols v. Raynbred* (1615) 80 Eng. Rep. 238, Hobart 88, repudiated by *Kingston v. Preston* (1773) 99 Eng. Rep. 437, 2 Douglas 689), there seems little sense in saying that the tenant must pay full rent even if the landlord has not performed his duties under the lease. Most would presumably agree with this, even if they disagree over what duties (if any) the law should imply into all leases.

Impact litigators played a part in bringing about the doctrinal shift in *Javins*. As Chused (2004) reports, the litigation was initiated by a legal services office created with new federal funds made available by President Johnson's War on Poverty. The impact litigators behaved very much as modeled: Their target was not an individual client but a change in landlord-tenant law, which would inure generally to the benefit of the urban poor. They selected a sympathetic lead plaintiff. And they underlined (then) legally irrelevant but practically salient facts that could sway judges, going so far as to bring "bags of mouse feces, dead mice, [and] roaches" from tenants' apartments to the courtroom (Chused (2004) 192, 211).

What separates the impact litigators in *Javins* from their more famous counterparts like the NAACP's Legal Defense Fund is the limited extent of their success. The paradigm shift from property law to contract was incomplete and pro-tenant reform went only so far, much to the dismay of left-leaning commentators (Merrill and Smith (2001); Kelley (1995)). This limited success was partly due to the fact that the pro-tenant impact litigators did not have unrivaled control over the selection of cases that made landlord-tenant law. In addition to their systematic selection, privately directed litigation ("random case selection" in the model) continued to shape general law (contrast the NAACP's impact litigation over racial segregation in education (Tushnet (1987))). For advocates of pro-tenant reform, the impact litigators' intervention was obviously welcome, albeit insufficiently influential. But even for those who do not share the impact litigators' vision of far-reaching leftward reform, their involvement contributed to rationalizing doctrine by importing some contract law principles into a field dominated by arcana of real property law. From this perspective, it was good that the impact litigators had some influence, but not too much influence, over case selection.

3.3 Hard Cases Make Good Law?

The models with under-inclusive doctrine assume that some inherent limitation makes doctrine incapable of reflecting certain kinds of case facts. The limitation's existence is important to the models (though its source is not). It implies that cases raising salient but extra-doctrinal factual issues create distractions that might divert judges from the path of true law. But imagine instead that cases with salient extra-doctrinal facts point up important issues that doctrine should but currently does not address. Then the doctrinal variations caused by cases posing a special hardship are not always distortions; they might be appropriate concessions to previously unheeded realities.

Consider *Winterbottom* in this light. A modern reader is not likely to commend the court for sticking to the privity-of-contract rule in spite of the coach driver's misfortune; to the contrary, the case shows the silliness of the rule. Indeed that is how courts eventually came to view the matter (e.g., *MacPherson v. Buick*, 217 N.Y. 382 (1916) (Cardozo, J.); *Donoghue v. Stevenson* (1932) UKHL 100; see also Shadmehr, Cameron and Shahshahani (2022 forthcoming)). A plausible interpretation is that confrontation with doctrinally unaccounted-for facts in cases like *Winterbottom* served in time to alert common law judges to the shortcomings of existing doctrine, causing them to revise outdated laws. In this interpretation, hard cases make good law by forcing the consideration of important unconsidered issues.

Of course, it is neither realistic nor theoretically interesting to conceptualize hard cases as having only this enlightening quality. A more promising approach would be to recognize that hard cases can distract judges from good lawmaking *or* teach judges something about the world. And instead of assuming that the dimensionality of doctrine is inherently limited, one can assume that expanding the dimensionality comes at a cost. The desirability of expansion would then depend on the distribution of case facts along the accounted- and unaccounted-for dimensions. In such a learning model, courts first update their priors about the distribution of global case facts by seeing new cases, and then make law. An impact litigator may help the court by providing accurate information about the distribution of case facts or mislead the court in order to locate the rule closer to her own ideal point, so the court may not update or may update skeptically. The strategic issues raised by such a setting are common to many signaling models (see Milgrom (2008) and Sobel (2009) for surveys). The extension can be profitably pursued in future work.

3.4 Equity Preserves Law

The model provokes us to see equity in a new light. The traditional justification for equitable doctrines is to prevent injustice and harshness to a party resulting from the strict application of law. That is how Roscoe Pound saw it when he spoke of equity as "a needed safety valve in the working of our legal system" (Pound (1922)). Smith (2012) offers a similar vision in his "reconstruction of the traditional approach to equity," which sees equity as frustrating sophisticated parties' "opportunism" in taking unfair advantage of the rigid structures of common law, a vision formalized in Ayotte, Friedman and Smith (2013).

The present analysis offers a different justification—to preserve sound law. In this view, equity affords an outlet to a court who would otherwise distort the law in order to achieve justice or fairness in a hard case. So the prime virtue of equity is not to save a particular party from injustice but to save many potential future parties (society at large) from the court's bending the law to save that party from injustice.

For example, consider laches—the equitable doctrine which holds that a claimant who waits unreasonably long before bringing suit is time-barred even if technically within the statute of limitations. The traditional view would say that laches protects the sympathetic defendant who has made decisions in reliance on the reasonable expectation that the claimant will not sue. The "hard cases" view would say that laches protects us all from the court's protecting the sympathetic defendant and denying relief to the claimant by distorting the applicable substantive or procedural law (e.g., the definition of when a claim accrues).

The difference in viewpoints comes down to a difference in assumptions about what the court would do in the absence of an equitable escape hatch. The traditional view assumes a farsighted court who would let the party in suit suffer a hardship rather than alter generally sound law. The present view, by contrast, allows that the court might be myopic and swayed by the party's plight to distort the law. Of course we don't know the precise distribution of farsighted and myopic judges, but the alternative view has purchase to the extent that "hard cases make bad law" has purchase.

Note well that for this view to hold water, equity must be relatively soft; it cannot be too law-like. If an equitable doctrine has definite preconditions for application, then there is no meaningful difference (analytically, in terms of the model) between equity and law. Sometimes the law is initially harsh and subsequently evolves to incorporate equitable considerations; but such situations are more usefully conceptualized in the dynamic terms of Section 3.3. By contrast, the idea of equity elaborated in this section is that the law *cannot* systematically incorporate certain considerations, so it allows for an external escape valve to keep doctrine pure. In this view, then, the lawlessness of equity is not a bug but an essential feature of its law-preserving function. The famous "Chancellor's foot" critique of equity (Selden (1689))[11] should be reinterpreted as praise.

This is not to say that equity is necessarily superior. The benefits of flexibility must be weighed against the costs of misapplication—both unintentional (because the boundaries of doctrine are unclear) and strategic (e.g., by a lower court with outlying preferences). This may be viewed as one manifestation of a fundamental tradeoff in principal-agent theory (e.g., Gilligan and Krehbiel (1987), Epstein and O'Halloran (1994), Aghion and Tirole (1997), Dessein (2002)). All the same, the point here is that the "safety valve" function of equity is not simply to attain justice in particular cases but, perhaps more importantly, to prevent the law from being distorted in order to attain justice in particular cases.

4 Conclusion

Through a series of formal models built around the maxim that hard cases make bad law, this paper sheds light on two fundamental features of the common law process. First, gen-

[11] "Equity is a roguish thing: for law we have a measure, know what to trust to; equity is according to the conscience of him that is Chancellor, and as that is larger or narrower, so is equity. 'Tis all one as if they should make the standard for the measure we call a foot, a Chancellor's foot; what an uncertain measure would this be? One Chancellor has a long foot, another a short foot, a third an indifferent foot: 'tis the same thing in a Chancellor's conscience."

eral laws are made by particular cases, so the vagaries of a few cases may matter a great deal beyond those cases. Second, strategic case selection by litigators attentive to judges' amenability to the influence of extra-doctrinal considerations can affect doctrine. The models capture the core qualitative intuition that a special hardship tempts courts to distort law. They also sharpen and enrich that intuition by considering how three different senses of "hard"—doctrinal under-inclusiveness (special hardship), case importance, and case difficulty (closeness)—interact to affect lawmaking. Cases posing a special hardship may make bad law, and they are more likely to do so when the case is important or close. Beyond the context assumed in the aphorism, the paper takes up impact litigators and judicial hierarchy. Surprisingly, impact litigators have a positive impact on lawmaking even when their preferences are maximally far from the social ideal—as long as their power over case selection is modest. Their impact is more nuanced when their selection power is enhanced; the analysis identifies three distinct regimes depending on litigator preferences. In a judicial hierarchy with asymmetric information and factfinding discretion, even cases that do not pose a special hardship may make bad law. Moreover, the cases that make bad law are intermediately difficult—not the hardest cases but not the easiest cases either. And bad laws are bad in the strong sense of being Pareto-dominated.

Future work can build on the present analysis by incorporating dynamic learning effects that allow hard cases to exert a positive as well as negative influence on law development (see §3.3). It can also explore judicial techniques for limiting the distortionary effects of hard cases, including: disposing of a case by unpublished opinion, distinguishing troublesome precedent by limiting it to its facts or introducing dubious distinctions, and, in the case of most apex courts, case selection through a discretionary docket. These tactics soften the rule-disposition tradeoff by avoiding the imperative that ensuring the just disposition of a particular case must come at the expense of distorting the law. But they come at substantial cost. For example, repeated use of nonprecedential opinions diminishes confidence in the courts, and introducing dubious distinctions intro doctrine avoids one distortion only by introducing another. The present models may also be extended by adding a new fact dimension to the judicial hierarchy game or introducing competition between impact litigators with divergent preferences. Finally, the question of how the particular instantiation of a policy problem influences the general solution to that problem may be explored in non-judicial settings such as administrative and legislative lawmaking. For example, the War Powers Act attempted in response to particular abuses by President Nixon to rework the general balance of congressional-executive authority in warmaking. More recently, gun control legislation has been tailored to mass shootings in schools, but more pervasive problems of gun violence have not inspired legislation. The idea that the particular might influence the general, and that the influence might be distortionary, is thus endemic in the politics of policymaking. Of course, there are important differences between judicial and nonjudicial lawmaking. Deeper understanding of the judicial setting can help us see how far the idea travels in other settings.

Appendix: Formal Statements and Proofs

Section 2.3. The proofs are straightforward and follow the main text.

Section 2.4.

First I solve backward to characterize the unique subgame-perfect Nash equilibrium of the game. Then I prove the statements in Proposition 2, which are properties of the equilibrium. Without loss of generality, assume $r_L \geq 1/2$; the proof for $r_L < 1/2$ is symmetric.

When a case is brought (in the first or second stage), C's optimal decision rule is

$$r = \begin{cases} x_1 & \text{if } \; x_1 > 1/2 \; \text{ and } \; x_2 > x_1 \; \text{ and } \; e > x_1 - 1/2 \\ x_1 & \text{if } \; x_1 < 1/2 \; \text{ and } \; x_2 < x_1 \; \text{ and } \; e > 1/2 - x_1 \\ 1/2 & \text{otherwise} \end{cases} \tag{A1}$$

So L's payoff from bringing a case is

$$U_L^2 = \begin{cases} x_1 - r_L & \text{if } \; x_1 < 1/2 \; \text{ and } \; x_2 < x_1 \; \text{ and } \; e > 1/2 - x_1 \\ x_1 - r_L & \text{if } \; x_1 > 1/2 \; \text{ and } \; x_2 > x_1 \; \text{ and } \; e > x_1 - 1/2 \; \text{ and } \; r_L \geq x_1 \\ r_L - x_1 & \text{if } \; x_1 > 1/2 \; \text{ and } \; x_2 > x_1 \; \text{ and } \; e > x_1 - 1/2 \; \text{ and } \; r_L < x_1 \\ 1/2 - r_L & \text{otherwise} \end{cases} \tag{A2}$$

By (A2), LC's expected utility from not bringing case and proceeding to the second stage is

$$\begin{aligned} U_L^2 &= \int_0^{0.5}\int_{0.5-x_1}^{0.5}\int_0^{x_1} x_1 - r_L dF_{X_2}(x_2)dF_E(e)dF_{X_1}(x_1) + \int_{0.5}^{r_L}\int_{x_1-0.5}^{0.5}\int_{x_1}^{1} x_1 - r_L dF_{X_2}(x_2)dF_E(e)dF_{X_1}(x_1) \\ &+ \int_{r_L}^{1}\int_{x_1-0.5}^{0.5}\int_{x_1}^{1} r_L - x_1 dF_{X_2}(x_2)dF_E(e)dF_{X_1}(x_1) + (5/6)(0.5-r_L) \\ &= -\frac{r_L^4}{3} + \frac{4r_L^3}{3} - 2r_L^2 + \frac{r_L}{3} + \frac{1}{6}. \end{aligned} \tag{A3}$$

At the first stage, L brings case iff $U_L^1 \geq U_L^2$. There are two ways this can happen: (1) $-|x_1 - r_L| \geq U_L^2$ (i.e., when the case makes bad law), (2) $-|1/2 - r_L| \geq U_L^2$ (i.e., when the case makes good law). Inequality (2) is satisfied for all $r_L \geq 1/2$, so L always brings a case that would make good law. For inequality (2), we must find the largest interval $[\underline{x}_1, \overline{x}_1]$ such that $x_1 \in [\underline{x}_1, \overline{x}_1] \implies -|x_1 - r_L| \geq U_L^2$. This yields $\underline{x}_1 = -\frac{r_L^4}{3} + \frac{4r_L^3}{3} - 2r_L^2 + \frac{4r_L}{3} + \frac{1}{6}$ and $\overline{x}_1 = \frac{r_L^4}{3} - \frac{4r_L^3}{3} + 2r_L^2 + \frac{2r_L}{3} - \frac{1}{6}$. It turns out that $\underline{x}_1 < 1/2\ \forall r_L \in (1/2, 1)$ (and $\underline{x}_1 = 1/2$ for $r_L = 1$). Setting $\overline{x}_1 = 1$ and solving for r_L yields an irrational number slightly smaller than 0.75, which we denote $\overline{r}_L$. Putting all this together, L's equilibrium strategy is given by the following decision rule:

$$L \text{ does } \textit{not} \text{ bring case iff} \begin{cases} x_1 < \underline{x}_1 \quad \text{and} \quad x_2 < x_1 \quad \text{and} \quad e > 1/2 - x_1 \\ \textit{or} \\ x_1 > \overline{x}_1 \quad \text{and} \quad x_2 > x_1 \quad \text{and} \quad e > x_1 - 1/2 \end{cases} \tag{A4}$$

where $\underline{x}_1 = -\frac{r_L^4}{3} + \frac{4r_L^3}{3} - 2r_L^2 + \frac{4r_L}{3} + \frac{1}{6}$ and $\overline{x}_1 = \begin{cases} \frac{r_L^4}{3} - \frac{4r_L^3}{3} + 2r_L^2 + \frac{2r_L}{3} - \frac{1}{6} & \text{if } r_L \leq \overline{r}_L \\ 1 & \text{if } r_L > \overline{r}_L \end{cases}$ and $\overline{r}_L$ is the value of r_L (an irrational number) that solves $\overline{x}_1(r_L) = 1$ ($\overline{r}_L \cong 0.749342$). Together, equations (A1) and (A4) characterize the unique subgame-perfect Nash equilibrium strategy profile of the game in this section. We are now in a position to prove Proposition 2.

Proof of Proposition 2.1. By (A1), expected welfare in the game without an impact litigator is

$$W_1 = \int_0^{0.5}\int_0^{x_1}\int_{0.5-x_1}^{0.5} x_1 - 0.5 dF_E(e)dF_{X_2}(x_2)dF_{X_1}(x_1) + \int_{0.5}^{1}\int_{x_1}^{1}\int_{x_1-0.5}^{0.5} 0.5 - x_1 dF_E(e)dF_{X_2}(x_2)dF_{X_1}(x_1) = -1/48. \tag{A5}$$

Based on (A1), (A4), and (A5), welfare in the game with an impact litigator is given by

$$W_2 = \begin{cases} x_1 - 1/2 & \text{if } x_1 \in (\underline{x}_1, 1/2) \text{ and } x_2 < x_1 \text{ and } e > 1/2 - x_1 \\ 1/2 - x_1 & \text{if } x_1 \in (1/2, \overline{x}_1) \text{ and } x_2 > x_1 \text{ and } e > x_1 - 1/2 \\ -1/48 & \text{if } L \text{ does not bring case in the first stage} \\ 0 & \text{otherwise} \end{cases} \tag{A6}$$

Based on (A6), expected welfare in the game with an impact litigator is calculated by integration to be

$W_2 = \frac{1}{15552}\left(-324 - 6(1+8r_L-12r_L^2+8r_L^3-2r_L^4)^4 - 23(-1-8r_L+12r_L^2-8r_L^3+2r_L^4)^3\right)$ (if $r_L \geq \overline{r}_L$)

and

$W_2 = \frac{1}{3888}(-1706 + 4941r_L + 9822r_L^2 - 39748r_L^3 - 6045r_L^4 + 134244r_L^5 - 175848r_L^6 + 116976r_L^7 - 125136r_L^8 + 205940r_L^9 - 230904r_L^{10} + 165840r_L^{11} - 79964r_L^{12} + 26304r_L^{13} - 5760r_L^{14} + 768r_L^{15} - 48r_L^{16})$ (if $r_L < \overline{r}_L$).

Using these expressions for W_2, one can verify that $W_2(r_L) > W_1 \; \forall r_L \in [1/2, 1]$ (see also Figure 5). □

Proof of Proposition 2.2. Using the same expressions for W_2, one can verify that $\partial W_2/\partial r_L < 0 \; \forall r_L \in [1/2, 1]$. □

Proof of Proposition 2.3. By (A4), the probability that L will *not* bring case in the first round (P) is calculated to be

$$P = \begin{cases} \frac{-1}{324}\left(-1-8r_L+12r_L^2-8r_L^3+2r_L^4\right)^3 & \text{if } r_L \geq \overline{r}_L \\ \\ \frac{1}{324}\left((1+8r_L-12r_L^2+8r_L^3-2r_L^4)^3 - (-7+4r_L+12r_L^2-8r_L^3+2r_L^4)^3\right) & \text{if } r_L < \overline{r}_L \end{cases} \tag{A7}$$

One can verify that the equation $\partial P/\partial r_L = 0$ has a unique solution in the interval

$r_L \in [1/2, 1]$ (denoted $\tilde{r}$, with $\tilde{r} \cong 0.718337$), with $\partial P/\partial r_L < 0$ for $r_L < \tilde{r}$ and $\partial P/\partial r_L > 0$ for $r_L < \tilde{r}$ (see also Figure 4). $\square$

Section 2.5.

Without loss of generality, assume $r_L \geq 1/2$. The proofs for $r_L < 1/2$ are symmetric. To avoid proliferation of subscripts, I abuse notation and use r instead of r_L for L's ideal rule.

Proof of Proposition 3.1. To begin note that

$D_1 = E(|r - X|) = E(r - X|X < r)\Pr(X < r) + E(X - r|X > r)\Pr(X > r) = r^2 - r + 1/2$.

Moreover, $D_{n+1} = E\{|X - r|\big|X \in B_{D_n}(r)\}\Pr(X \in B_{D_n}(r)) + D_n(1 - \Pr(X \in B_{D_n}(r)))$. So (D_n) is decreasing. It is easy to verify that $r - D_1 < 1/2\,\forall\, r < 1$. There are two cases to consider: (1) $r + D_1 \leq 1$ and (2) $r + D_1 > 1$, which is to say (1) $r \leq 1/\sqrt{2}$ and (2) $r > 1/\sqrt{2}$. For case (1), note that $D_{n+1} = D_n(1 - D_n)$. To see that $(D_n) \to 0$, suppose for contradiction that (D_n) does not converge to 0. Then, given that (D_n) is bounded below by 0, there exists $\epsilon \in (0, 1)$ such that $D_n > \epsilon\,\forall\, n$. Now $D_n > \epsilon \implies 1 - D_n < 1 - \epsilon$, so for all n we have

$D_{n+1} = D_n(1 - D_n) < D_n(1 - \epsilon) < D_{n-1}(1 - \epsilon)^2 < ... < D_1(1 - \epsilon)^n$. But $\lim_{n\to\infty} D_1(1 - \epsilon)^n = 0$, which implies that $(D_n) \to 0$, a contradiction.

For case (2), if $r = 1$ then

$D_{n+1} = E(1 - X|X > 1 - D_n)\Pr(X > 1 - D_n) + D_n \Pr(X < 1 - D_n) = D_n(1 - D_n/2)$, and it follows by the same argument as in case (1) that the sequence converges to 0.

Finally consider the case $r \in (1/\sqrt{2}, 1)$. As before,

$D_{n+1} = E\{|X - r|\big|X \in B_{D_n}(r)\}\Pr(X \in B_{D_n}(r)) + D_n(1 - \Pr(X \in B_{D_n}(r)))$, so (D_n) is decreasing. Now either there exists $\overline{n}$ such that $r + D_n < 1\,\forall n \geq \overline{n}$ or there does not exist such $\overline{n}$. If such $\overline{n}$ exists then for all $n \geq \overline{n}$ the sequence takes the form $D_{n+1} = D_n(1 - D_n)$, and $(D_n) \to 0$ by the same argument as in case (1). If such $\overline{n}$ does not exist (which implies that (D_n) does not converge to 0) then for all n we have

$D_{n+1} = E(X - r|X \in (r, 1))\Pr(X \in (r, 1)) + E(r - X|X \in (r - D_n, r))\Pr(X \in (r - D_n, r)) + D_n \Pr(X < r - D_n) = \frac{(1 - r)^2}{2} + D_n(r - \frac{D_n}{2})$. Now recall the sequence we obtained when $r = 1$ and call that sequence (D'_n) (with $D'_{n+1} = D'_n(1 - D'_n/2)$). Note that $D'_1 > D_1 = r^2 - r + 1/2\,\forall\, r < 1$, and note that one can verify (using the assumption that $r + D_n > 1\,\forall\, n$) that if there exists some n such that $D'_n > D_n$ then $D'_{n+1} > D_{n+1}$. It follows that $D_n < D'_n\,\forall\, n$, which implies, because $(D'_n) \to 0$, that $(D_n) \to 0$, proving our claim and contradicting the assumption that the hypothesized $\overline{n}$ does not exist. $\square$

Proof of Proposition 3.2. Recall that $D_1 = r^2 - r + 1/2$. Accordingly $DW_1 = 1/4$. Now $DW_2 = E(|X - 1/2| \big| X \in B_{D_1}(r)) \Pr(X \in B_{D_1}(r)) + (1/4)(1 - \Pr(X \in B_{D_1}(r)))$. Note that $E(|X - 1/2| \big| X \in B_{D_1}(r)) < 1/4 \, \forall r < 1$ (and $E(.) = 1/4$ for $r = 1$), so $DW_2 < DW_1 \, \forall r < 1$ (and $DW_2 = DW_1$ for $r = 1$). $\square$

Proof of Propositions 3.3-3.5. We begin with a useful lemma.

Lemma 1. *For any n such that $B_{D_n}(r) \in [1/2, 1]$,*
$DW_n > DW_{n+1} > DW_{n+2} > ...$ *if* $DW_n > r - 1/2$ *and*
$DW_n < DW_{n+1} < DW_{n+2} < ...$ *if* $DW_n < r - 1/2$.

Proof of Lemma 1. Note that $DW_1 = 1/4$ and, for all n,
$DW_{n+1} = E\{|X - 1/2| \big| X \in B_{D_n}(r)\} \Pr(X \in B_{D_n}(r)) + DW_n(1 - \Pr(X \in B_{D_n}(r)))$. Now if $B_{D_n}(r) \in [1/2, 1]$ (which also implies by the decreasingness of (D_n) that $B_{D_{n+1}}(r) \in [1/2, 1]$) then we obtain $DW_{n+1} = (r - 1/2)2D_n + DW_n(1 - 2D_n)$. Because DW_{n+1} is a convex combination of $r - 1/2$ and DW_n, it follows that $DW_{n+1} > DW_n$ if $DW_n < r - 1/2$ and $DW_{n+1} < DW_n$ if $DW_n > r - 1/2$. $\square$

Now consider three cases separately: (1) $r \in (1/2, 1/\sqrt{2})$, (2) $r > 3/4$, (3) $r \in [1/\sqrt{2}, 3/4]$.

Case 1 ($r < 1/\sqrt{2}$): In this case $B_{D_n}(r) \in (0, 1) \, \forall n$, so
$DW_{n+1} = E\{|X - 1/2| \big| X \in B_{D_n}(r)\} 2D_n + DW_n(1 - 2D_n)$. Now for all n such that $r - D_n < 1/2$, we know that $E\{|X - 1/2| \big| X \in B_{D_n}(r)\}$ is increasing in D_n and therefore decreasing in n. So for all n such that $r - D_{n-1} < 1/2$, we have $DW_1 > DW_2 > ... > DW_n$. Note moreover that, for all n such that $r - D_n < 1/2$ and $B_{D_n}(r) \in (0, 1)$, we have $E\{|X - 1/2| \big| X \in B_{D_n}(r)\} > r - 1/2$. Now let n' be the first n such that $B_{D_n}(r) \in [1/2, 1]$. Because $E\{|X - 1/2| \big| X \in B_{D_{n'-1}}(r)\} > r - 1/2$ and $DW_{n'-1} > E\{|X - 1/2| \big| X \in B_{D_{n'-1}}(r)\}$, it follows that $DW_{n'} > r - 1/2$ and
$DW_{n'+1} = (r - 1/2)2D_{n'} + DW_{n'}(1 - 2D_{n'}) \in (r - 1/2, DW_{n'})$. Therefore, by Lemma 1, $DW_{n'+1} > DW_{n'+2} > ...$ We have shown (DW_n) is decreasing, which proves Proposition 3.3.

Case 2 ($r > 3/4$): Let n' be the first n such that $r - D_n \geq 1/2$. Then
$r - 1/2 > 1/4 = DW_1 > DW_2 > ... > DW_{n'} =$
$E\{|X - 1/2| \big| X \in B_{D'_{n-1}}(r)\} \Pr(X \in B_{D'_{n-1}}(r)) + DW_{n'-1}(1 - \Pr(X \in B_{D'_{n-1}}(r))$. Now

$DW_{n'+1} =$

$$\begin{cases} E\{X-1/2|X\in(r-D_{n'},1)\}\,\Pr(X\in(r-D_{n'},1))+DW_{n'}(1-\Pr(X\in(r-D_{n'},1)) & \text{if } \; r+D_{n'}\geq 1 \\ (r-1/2)2D_{n'}+DW_{n'}(1-2D_{n'}) & \text{if } \; r+D_{n'}<1 \end{cases}$$

which, in either case, yields $DW_{n'+1} > DW_{n'}$.

As for $DW_{n'+2}$: If $r + D_{n'} \leq 1$ then $r + D_{n+1} < 1 \,\forall\, n \geq n'$ and, because $DW_{n'+1} < r - 1/2$, it follows by Lemma 1 that $DW_{n'+1} < DW_{n'+2} < ...$ If $r + D_{n'} > 1$: If $r + D_{n'+1} \leq 1$ then, by the same argument as above, $DW_{n'+1} < DW_{n'+2} < ...$; if $r + D_{n'+1} > 1$ then $DW_{n'+2} =$ $E\{X-1/2\big|X\in(r-D_{n'+1},1)\}\,\Pr(X\in(r-D_{n'+1},1))+DW_{n'+1}(1-\Pr(X\in(r-D_{n'+1},1)))>DW_{n'+1}$ where the inequality follows from the fact that $E\{X-1/2\big|X\in(r-D_{n'+1},1)\} > E\{X-1/2\big|X\in(r-D_{n'},1)\} > DW_{n'+1}$. So we have $DW_{n'+1} < DW_{n'+2}$ and, repeating the same argument, we obtain $DW_{n'+2} < DW_{n'+3} < ...$ We have shown that, for all $r > 3/4$, there exists n' such that $DW_1 > DW_2 > ... > DW_{n'}$ and $DW_{n'} < DW_{n'+1} < ...$ Finally note that $(D_n) \to 0 \implies (DW_n) \to r - 1/2 > DW_1$, which concludes the proof of Proposition 3.5.

Case 3 ($r \in [1/\sqrt{2}, 3/4]$): Let n' be the first n such that $r - D_n \geq 1/2$ and note, for all $r \in [1/\sqrt{2}, 3/4]$), that $r - D_n \geq 1/2 \implies r + D_n \leq 1$. We know that $1/4 = DW_1 > ... > DW_{n'} = E\{|X-1/2|\big|X\in B_{D_{n'-1}}(r)\}\,\Pr(X\in B_{D_{n'-1}}(r))+DW_{n'-1}(1-Pr(.))$. Indeed, it turns out that $n' = 2\,\forall\, r \in [1/\sqrt{2}, 3/4]$ (and that $r - D_2 = 1/2$ for $r = 1/\sqrt{2}$). So $DW_{n'+1} = DW_3 = (r-1/2)2D_2 + DW_2(1-2D_2)$. One can calculate that $DW_2 = \dfrac{(1/2-r+D_1)^2}{2} + \dfrac{1}{8} + \dfrac{r-D_1}{4}$, and $DW_2 < r - 1/2 \iff r > \overline{r}$ where $\overline{r} \in (1/\sqrt{2}, 3/4)$ is an irrational number ($\overline{r} \approx 0.7349$).

$$\text{Therefore, } \begin{cases} DW_1 > DW_2 > ... & \text{for } \; r\in[1/\sqrt{2},\overline{r}) \\ DW_1 > DW_2 \;\text{ and }\; DW_2 < DW_3 < ... & \text{for } \; r\in(\overline{r},3/4] \\ DW_1 > DW_2 = DW_3 = ... & \text{for } \; r=\overline{r} \end{cases}$$

Note that $r \in [1/\sqrt{2}, \overline{r})$ fall in Proposition 3.3 and $r \in [\overline{r}, 3/4]$ fall in Proposition 3.4. This concludes the proof of Proposition 3. □

References

Aghion, Philippe and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy* 105(1):1–29.

Ayotte, Kenneth, Ezra Friedman and Henry E. Smith. 2013. "A Safety Valve Model of Equity as Anti-Opportunism.".
URL: *https://ssrn.com/abstract=2245098*

Badawi, Adam and Scott Baker. 2015. "Appellate Lawmaking in a Judicial Hierarchy." *Journal of Law & Economics* 58(1):139–172.

Baker, Scott and Claudio Mezzetti. 2012. "A Theory of Rational Jurisprudence." *Journal of Political Economy* 120(3):513–551.

Baker, Scott and Gary Biglaiser. 2014. "A Model of Cause Lawyering." *Journal of Legal Studies* 43(1):37–63.

Beim, Deborah, Alexander V. Hirsch and Jonathan P. Kastellec. 2014. "Whistleblowing and Compliance in the Judicial Hierarchy." *American Journal of Political Science* 58(4):904–918.

Cameron, Charles M., Jeffrey A. Segal and Donald Songer. 2000. "Strategic Auditing in a Political Hierarchy: An Informational Model of the Supreme Court's Certiorari Decisions." *American Political Science Review* 94(1):101–116.

Carrubba, Clifford J. and Tom S. Clark. 2012. "Rule Creation in a Political Hierarchy." *American Political Science Review* 106(3):622–643.

Chused, Richard. 2004. "Saunders (a.k.a. Javins) v. First National Realty Corporation." *Georgetown Law Journal on Poverty Law & Policy* 11:191–247.

Cooter, Robert and Lewis Kornhauser. 1980. "Can Litigation Improve the Law without the Help of Judges?" *Journal of Legal Studies* 9(1):139–163.

Cooter, Robert, Lewis Kornhauser and David Lane. 1979. "Liability Rules, Limited Information, and the Role of Precedent." *Bell Journal of Economics* 10(1):366–373.

Dessein, Wouter. 2002. "Authority and Communication in Organizations." *Review of Economic Studies* 69:811–838.

Dworkin, Ronald. 1975. "Hard Cases." *Harvard Law Review* 88:1057–1109.

Epp, Charles R. 1998. *The Rights Revolution: Lawyers, Activists, and Supreme Courts in Comparative Perspective*. University of Chicago Press.

Epstein, David and Sharyn O'Halloran. 1994. "Administrative Procedures, Information, and Agency Discretion." *American Journal of Political Science* 38(3):697–722.

Garner, Bryan. 2011. *Garner's Dictionary of Legal Usage.* Oxford University Press.

Gennaioli, Nicola and Andrei Shleifer. 2007. "The Evolution of Common Law." *Journal of Political Economy* 115(1):43–68.

Gilligan, Thomas W. and Keith Krehbiel. 1987. "Collective Decisionmaking and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics, & Organization* 3(2):287–335.

Hartocollis, Anemona. 2017. "A One-Man Legal Factory Fights Harvard Over Affirmative Action." *New York Times* Nov. 20, 2017:A1.
URL: *https://www.nytimes.com/2017/11/19/us/affirmative-action-lawsuits.html*

Heuston, R.F.V. 1978. "Hard Cases Make Bad Law." *Dublin University Law Journal* 2:31–35.

Kastellec, John P. 2017. The Judicial Hierarchy. In *Oxford Research Encyclopedia of Politics*. Oxford University Press.

Kelley, Robert. 1995. "Any Reports of the Death of the Property Law Paradigm for Leases Have Been Greatly Exaggerated." *Wayne Law Review* 41:1563–1607.

Kornhauser, Lewis. 1980. "A Guide to the Perplexed Claims of Efficiency in the Law." *Hofstra Law Review* 8(3):591–639.

Kornhauser, Lewis A. 1992. “Modeling Collegial Courts. II. Legal Doctrine.” *Journal of Law, Economics, & Organization* 8(3):441–470.

Lax, Jeffrey R. 2003. “Certiorari and Compliance in the Judicial Hierarchy: Discretion, Reputation and the Rule of Four.” *Journal of Theoretical Politics* 15(1):61–86.

Lax, Jeffrey R. 2011. “The New Judicial Politics of Legal Doctrine.” *Annual Review of Political Science* 14(1):131–157.

Lax, Jeffrey R. 2012. “Political Constraints on Legal Doctrine: How Hierarchy Shapes the Law.” *Journal of Politics* 74(3):765–781.

Mack, Kenneth W. 2012. *Representing the Race: The Creation of the Civil Rights Lawyer*. Harvard University Press.

Merrill, Thomas and Henry Smith. 2001. “The Property/Contract Interface.” *Columbia Law Review* 101:773–852.

Milgrom, Paul. 2008. “What the Seller Won’t Tell You: Persuasion and Disclosure in Markets.” *Journal of Economic Perspectives* 22(2):115–131.

Posner, Richard. 1973. *Economic Analysis of Law*. Little Brown.

Posner, Richard. 2002. *The Problematics of Moral and Legal Theory*. Belknapp Press.

Pound, Roscoe. 1922. *An Introduction to the Philosophy of Law*. Yale University Press.

Priest, George. 1977. “The Common Law Process and the Selection of Efficient Rules.” *Journal of Legal Studies* 6(1):65–82.

Rabin, Edward. 1984. “The Revolution in Residential Landlord-Tenant Law: Causes and Consequences.” *Cornell Law Review* 69:517–84.

Radin, Max. 1938. *The Law and Mr. Smith*. Bobbs Merrill.

Rubin, Paul. 1977. “Why Is the Common Law Efficient?” *Journal of Legal Studies* 6(1):51–63.

Schauer, Frederick. 2006. "Do Cases Make Bad Law." *University of Chicago Law Review* 73:883–918.

Selden, John. 1689. *The Table-Talk of John Selden*. Cambridge University Press (republished 2015).

Shadmehr, Mehdi, Charles Cameron and Sepehr Shahshahani. 2022 forthcoming. "Coordination and Innovation in Judiciaries: Correct Law vs. Consistent Law." *Quarterly Journal of Political Science* .
URL: *https://ssrn.com/abstract=3148866*

Shahshahani, Sepehr. 2021. "The Fact-Law Distinction: Strategic Factfinding and Lawmaking in a Judicial Hierarchy." *Journal of Law, Economics, & Organization* 37(2):440–477.

Shapiro, Scott J. 2007. The "Hart-Dworkin" Debate: A Short Guide for the Perplexed. In *Ronald Dworkin*, ed. Arthur Ripstein. Cambridge University Press.

Smith, Henry E. 2012. "The Equitable Dimension of Contract." *Suffolk University Law Review* 45:897–914.

Sobel, Joel. 2009. Signaling Games. In *Encyclopedia of Complexity and Systems Science*. Springer.

Sutton, Jeffrey. 2010. "A Review of Richard A. Posner, How Judges Think (2008)." *Michigan Law Review* 108:859–876.

Teles, Steven M. 2008. *The Rise of the Conservative Legal Movement: The Battle for Control of the Law*. Princeton University Press.

Tushnet, Mark V. 1987. *The NAACP's Legal Strategy Against Segregated Education, 1925–1950*. University of North Carolina Press.

Weinrib, Laura. 2016. *The Taming of Free Speech: America's Civil Liberties Compromise*. Harvard University Press.

Online Appendix: Judicial Hierarchy

The aphorism implicitly presupposes a single court making law by deciding a case, but in fact judiciaries are hierarchically structured. The literature on strategic interaction in the judicial hierarchy is vast (see Kastellec (2017) for a review). Going back to the seminal work of Cameron, Segal and Songer (2000), most of this literature focuses on auditing—that is, whether the higher court will review the lower court's decision. For example, Lax (2003) studies the impact of the "rule of four" on Supreme Court certiorari decisions; Beim, Hirsch and Kastellec (2014) examine how "whistle blowing" by a judge or a nonjudicial actor can help higher courts induce lower courts' compliance; and Badawi and Baker (2015) investigate an appellate court's investments in developing precedent in the context of auditing trial courts. The auditing framework is appropriate for studying a supreme court's review of appellate court decisions (as well as *en banc* review in federal circuit courts), where the decision whether to hear an appeal is discretionary, but not so much for intermediate appellate courts' review of trial court decisions, where the appellate court is required to hear all appeals. This section focuses on the latter (more common) level of hierarchical interaction.

A salient feature of this setting is that trial courts are better positioned to observe case facts, and appellate courts defer to trial courts' factual determinations under the "clear error" standard of review. Unlike trial courts' legal determinations, which can be overturned whenever the appellate court finds them to be wrong, trial courts' factual determinations cannot be overturned unless they are *clearly* wrong (see Shahshahani (2021), 441, for fuller exposition). The goal of this section is to understand how informational asymmetry and trial courts' fact discretion, enshrined in the clear-error standard, moderate the influence of case characteristics on lawmaking.

Model

The model is based on Shahshahani (2021). The players are a higher court HC, with ideal rule H, and a lower court LC, with ideal rule L. Ideal rules are common knowledge. The fact space is one-dimensional and doctrine is perfectly inclusive. Sequence of play is as follows:

1. Nature selects the true case facts ($x_t \in \mathbb{R}$) and a signal of case facts ($x \in \mathbb{R}$). LC observes both x_t and x, but HC observes only x. From HC's perspective, true case facts are uniformly distributed on an epsilon ball around the signal: $(X_t \,|\, X = x) \sim U[x - \epsilon, x + \epsilon] \; \forall x$.

2. LC decides what facts to report, x', where $x' \in [x - \epsilon, x + \epsilon]$.

3. HC announces the rule r, which determines the disposition as follows: $d = \begin{cases} 1 & \text{if } x' < r \\ 0 & \text{if } x' > r \end{cases}$.

   If $r = x'$ then HC can choose either disposition.

Payoffs are:

$$U_{HC} = -|r - H| + e_h \mathbb{1}\{d = d_H\} \tag{OA1}$$

$$U_{LC} = -|r - L| + e_\ell \mathbb{1}\{d = d_L\} \tag{OA2}$$

where d_L and d_H are the courts' ideal dispositions, as in equation (2).

A strategy for LC is the choice of facts to report in light of the public signal and the true facts $\big(\sigma_{LC} : \mathbb{R} \times [x - \epsilon, x + \epsilon] \to [x - \epsilon, x + \epsilon]\big)$. A strategy for HC is the choice of a rule in light of the public signal and LC's reported facts $\big(\sigma_{HC} : \mathbb{R} \times [x - \epsilon, x + \epsilon] \to \mathbb{R}\big)$. The solution concept is perfect Bayesian equilibrium. Without loss of generality, assume $L > H$ and $H = 0$.

Analysis

Judicial preferences in this model are like the single-court models: Courts are interested both in getting the right disposition in this case and in making the right rule to govern future cases (though the two courts don't have the same view of what constitutes the "right" rule and disposition). However, the informational environment is different from the single-court context in that the trial court knows more about case facts than the appellate court. The appellate court knows the neighborhood of true case facts whereas the trial court knows the precise location. (Of course, x_t need not be the literal truth; it can be a best estimate.) The parameter ϵ indexes the radius of the neighborhood. A larger ϵ denotes a more fact-intensive case, such that the public signal conveys only a general indication of where the true facts are and the trial court's factfinding discretion is concomitantly greater.

The model gives content to the deferential clear-error standard of review by assuming that as long as the trial court's reported facts are within the epsilon neighborhood where the truth could be ($x' \in [x - \epsilon, x + \epsilon]$), they are not clearly erroneous and must be accepted by the appellate court.[1] So the appellate court's rule generates a disposition by reference to the case facts reported by the trial court (x'), not the public signal (nor the true facts).

The source of strategic tension in the model is the trial court's use of its factfinding discretion to obtain its preferred disposition. In particular, when the two courts' ideal dispositions conflict, the trial court is tempted to slant the facts to get the disposition it wants. On the other hand, the appellate court understands this strategic incentive, so it may not believe the facts reported by the trial court. And even though it cannot directly override those facts (because of the clear-error standard of review), it can distort the rule to change the disposition. So, as in Section 2.3, the appellate court faces a rule-disposition tradeoff, but the source of the tradeoff is different and it exists even when doctrine is perfectly inclusive.

Moreover, the appellate court's Bayesian assessment of the trial court's factfinding is complicated by the dual nature of such factfinding. When the trial court reports facts other than the public signal ($x' \neq x$), it may be misrepresenting the true facts; on the other hand, it may be attempting to correct the mistaken impression of true facts created by the signal (see Figure OA1). It is useful to distinguish these two varieties of factfinding. "Helpful factfinding" occurs when the trial court uses its factfinding power to report facts that are on the same side of the appellate court's ideal point as the true facts; "deceptive factfinding" occurs when the trial court uses its factfinding power to report facts that are on the opposite side of the appellate court's ideal point as the true facts. Formally, helpful factfinding means $x' \neq x$ and $\text{sign}\{x'\} = \text{sign}\{x_t\}$; deceptive factfinding means $x' \neq x$ and $\text{sign}\{x'\} \neq \text{sign}\{x_t\}$.

With a sense of the strategic forces at play, we are now in a position to discuss the model's equilibrium. The first important result is that very easy cases never make bad law.

[1] In the model the trial court is restricted to finding facts within an epsilon neighborhood of the public signal. One can construct a more complicated model where the trial court can report facts anywhere but reported facts outside the epsilon neighborhood are clearly erroneous and reversible. But this would reduce to essentially the same model because reporting clearly erroneous facts would be dominated.

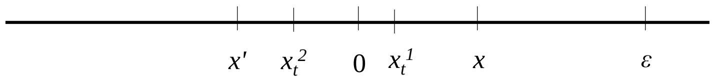


Figure OA1: HC sees the signal (x) and LC's factual report (x'), but not the true facts (x_t), so it does not know whether LC's factfinding is helpful or deceptive. If $x_t < 0$ (as in x_t^2) then the factfinding is helpful; but if $x_t > 0$ (as in x_t^1) then the factfinding is deceptive.

If the public signal is very far from the appellate court's ideal rule (i.e., if $x > \epsilon$ or $x < -\epsilon$), then the appellate court, though uncertain about the precise location of case facts, knows all that it needs to know. If the public signal is very far to the left then the appellate court knows that the true facts are also to the left of its ideal point ($x < -\epsilon \implies x_t < 0$), so its ideal disposition is 1. By the same token, if the public signal is very far to the right then its ideal disposition is 0. Moreover, when the public signal is so extreme, the trial court cannot move the operative facts from one side of the appellate court's ideal rule to the other (i.e., $\text{sign}\{x\} = \text{sign}\{x'\}$). Therefore, for extreme public signals (very easy cases), setting the rule at its ideal point is the unique maximizer of the appellate court's payoff function—regardless of the facts reported by the trial court.[2]

Remark 2. *In the judicial-hierarchy game with perfectly inclusive doctrine, very easy cases never make bad law. Formally,* $x \notin [-\epsilon, \epsilon] \implies r = 0$.

Note well the generality of this result. Remark 2 does not pertain to a specific equilibrium of the game; rather, it says there exists no equilibrium in which very easy cases ever make bad law. It turns out that similarly general results cannot be stated about case importance. But considering an equilibrium of the game where bad lawmaking can happen shows how judicial hierarchy complicates the analysis from a single-court model.

Figure OA2 portrays the equilibrium outcomes of factfinding and rulemaking as a function of the public signal (see Proposition 5 for a full characterization of perfect Bayesian equilibrium). To understand the salient features of this equilibrium, it is useful to consider different realizations of the public signal in turn. For extreme realizations of the public signal ($x \notin (-\epsilon, \epsilon)$), HC always sets the rule at its ideal point (see Remark 2).

Next consider the region $x \in (0, \epsilon)$. It is useful to consider what would happen if HC were to always set its ideal rule ($r = 0$). Then, if $x_t > L$ then LC would not have an incentive

[2]The foregoing, and the discussion of equilibrium to follow, take the location of x to be the measure of difficulty. Given the distribution of $X_t \mid X$, similar results would obtain if x_t were taken to be the measure.

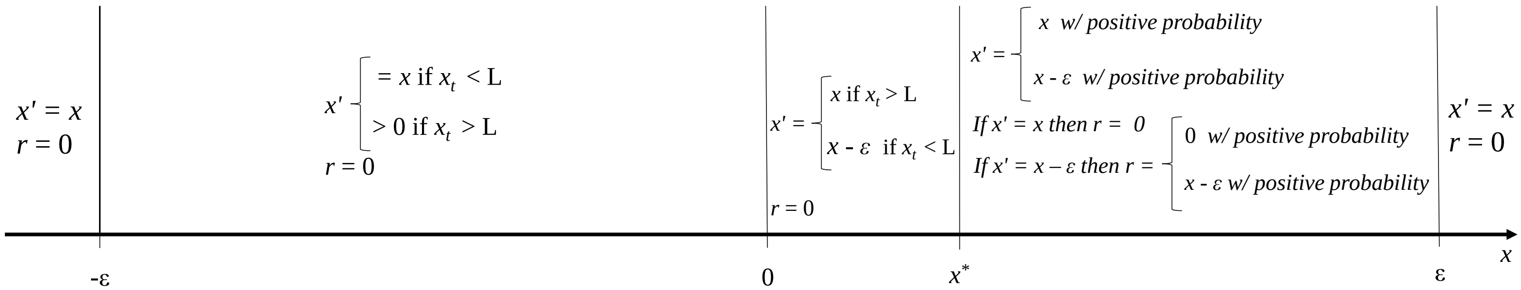


Figure OA2: Equilibrium outcomes of factfinding and rulemaking as a function of the public signal in the judicial-hierarchy game with perfectly inclusive doctrine. Cases make bad law with positive probability when $x \in (x^*, \epsilon)$.

to report facts other than the public signal because it could get its preferred disposition under the public signal. However, if $x_t < L$ then LC would set $x' < 0$ to get its preferred disposition, which it could not get without such a fact report. Now consider whether HC would keep the rule at $r = 0$ in response. As discussed in connection with Figure OA1, HC's decision is complicated by the fact that it does not know whether LC's factfinding is helpful or deceptive. If LC's factfinding is helpful then HC is better off keeping the rule at 0 because that would produce both its preferred rule and its preferred disposition. But if LC's factfinding is deceptive then HC might be better off setting $r = x'$ in order to reverse the dispositional effect of the factfinding and prevent the loss of dispositional utility—provided, however, that the gain in dispositional utility is worth the cost in rule utility that would be borne by setting $r = x'$. (Note that HC would never distort the rule away from its ideal point more than the minimum extent necessary to reverse the dispositional effect of LC's factfinding. So, when $x' < 0$, HC would never set $r < x'$.) Ultimately, then, HC's response to LC's reporting facts $x' < 0$ depends on two considerations: (1) HC's posterior belief that LC's factfinding is deceptive, and (2) the amount of rule utility that HC would have to sacrifice to guard against the probable loss of dispositional utility. For values of x close to 0, both considerations lead HC toward keeping the rule at its ideal point; for values of x close to ϵ, by contrast, both considerations pull HC toward choosing $r = x'$ in response to $x' < 0$.

These dynamics lead to an equilibrium with a threshold structure. The threshold x^* in the interval $(0, \epsilon)$ specifies the value of x at which, provided LC sets x' as far to the left of 0 as possible (which is in its interest to do), HC's expected utilities from $r = 0$ and $r = x'$ are equal. HC would "tolerate" LC's factfinding where the public signal is below x^* but not above x^*—meaning that if $x > x^*$ and $x' < 0$ then HC would set $r = x'$ to counteract the factfinding. (See Proposition 5 for the closed form of x^*.)

LC's factfinding in turn depends on HC's anticipated response. Below x^*, HC would tolerate fact reports far to the left of 0, so LC reports facts other than the public signal if and only if doing so would be necessary to obtain its preferred disposition (i.e., if $x_t > L$ then $x' = x$, and if $x_t < L$ then $x' = x - \epsilon$). Above x^*, as discussed, HC would not tolerate LC's strategy of engaging in factfinding whenever necessary to obtain LC's preferred disposition. Rather, in equilibrium, LC sometimes engages in factfinding and HC sometimes tolerates it. The key to sustaining this strategy profile in equilibrium is that LC is more likely to engage in helpful than deceptive factfinding. (That is, $\Pr(x' < 0|x_t < 0) > \Pr(x' < 0|x_t \geq 0)$. The precise relationship between the probabilities is stated in Proposition 5.) In the absence of this relationship between the probabilities of helpful and deceptive factfinding, HC would never tolerate factfinding above x^*. As it is, HC sometimes tolerates LC's factfinding ($r = 0$) and sometimes does not ($r = x'$), and the choice not to tolerate is what produces bad laws.

Finally, consider the region $x \in (-\epsilon, 0)$. Again it is useful to consider what would happen if HC were to always set its ideal rule. Then, again, LC would report facts other than the public signal if and only if doing so would be necessary to obtain its ideal disposition (i.e., $x' > 0$ if $x_t \geq L$, and $x' = x$ if $x_t < L$).[3] But here, unlike when $x \in (0, \epsilon)$, such factfinding can never be deceptive because if $x_t \geq L$ then $x_t > 0$ as well. Therefore, HC always keeps the rule at its ideal point and the case never makes bad law.

Having worked through the equilibrium logic, let us step back and consider the implications for the substantive questions motivating the analysis. The most general lesson is that in a judicial hierarchy with factfinding discretion, unlike in the single-court context, cases may make bad law even if doctrine is perfectly inclusive (contrast Remark 1). The intuition is that when the trial court is more informed about case facts than the appellate court, its strategic factfinding confronts the appellate court with a rule-disposition tradeoff. When that tradeoff is finely balanced, the appellate court randomizes between sacrificing the disposition to the rule and sacrificing the rule to the disposition, and the latter choice produces

[3] In this region, factfinding beyond the public signal can occur with positive probability only if the two courts' ideal points are close (namely, if $L < \epsilon$). If $L \geq \epsilon$ then the fact that $x < 0$ implies $x_t < L$, so LC would always get its preferred disposition under the public signal.

bad laws. Rule distortion, then, is the result of the trial court's strategic exploitation of the appellate court's uncertainty and the latter's deferential review of factfinding.

Note that strategic interaction in the judicial hierarchy is necessary for this result; introducing uncertainty about case facts into the single-court context would not have produced bad laws. In the single-court context with perfectly inclusive doctrine, even if the court receives an arbitrarily noisy signal of case facts, it would still always announce the ideal rule as long as its signal is accurate in expectation. This follows immediately from expected-utility maximization and the logic of Section 2.2. In that context, some cases would get the wrong disposition, but no case would make bad law.

The next important result is that the effect of case difficulty (parametrized by x) on the quality of lawmaking is nonmonotonic.[4] The cases that may make bad law are not the easiest cases ($x > \epsilon$), nor the most difficult cases ($x \in (0, x^*)$), but rather intermediately difficult cases ($x \in (x^*, \epsilon)$). When a case is really easy, the public signal accurately conveys the location of true facts in relation to the appellate court's ideal rule, and the trial court's factfinding cannot change that, so the appellate court never distorts the rule (see Remark 2 and discussion). When a case is really difficult, (1) the equilibrium probability that the trial court's factfinding is deceptive is relatively low, and (2) the appellate court would have to sacrifice a great deal of rule utility to counteract the dispositional effect of the factfinding, so the court does not distort the rule. Instead, rule distortion occurs in the intermediate-difficulty range where the probability of deceptive factfinding is relatively high and a probably-bad disposition can be prevented by a modest sacrifice in rule utility.

Unlike case difficulty (Remark 2), there is no nonzero level of case importance to the trial or appellate court that would guarantee good lawmaking. For any nonzero value of e_h or e_ℓ, there are regions in the parameter space where the rule is distorted with positive probability.

The final lesson of this section is that when a judicial hierarchy makes bad laws, it makes laws that are bad for *both* courts. The distorted rule is bad not just in the sense of deviating from the higher court's ideal rule but also in the strong sense of being Pareto-dominated.

[4] The effect is also asymmetric. When the public signal is to the left of the appellate court's ideal rule ($x < 0$), bad lawmaking never happens, regardless of case difficulty. The intuition, again, is that in this region the trial court's factfinding to help itself also helps the appellate court. By contrast, when the public signal is to the right of the appellate court's ideal rule ($x > 0$), bad law may be made. Of course, the asymmetry would be on the opposite side if $L < 0$.

That is, there are rules that both courts would prefer to the distorted rule. (Namely, when the rule is distorted, we have $r < 0$, so there are laws in the interval $[0, L]$, e.g., $r = 0$, that both courts would prefer.) This strong form of distortion is required to reverse the dispositional effect of the trial court's factfinding and to punish the trial court. A rule on the Pareto frontier ($r \in [0, L]$) would fail to accomplish both tasks.

Proposition 4 summarizes the insights from the judicial-hierarchy game.

Proposition 4. *In the judicial-hierarchy model with perfectly inclusive doctrine,*

1. *Unlike in the single-court context with perfectly inclusive doctrine, some cases may make bad law.*
2. *Very easy cases never make bad law.*
3. *There is no nonzero level of case importance to the trial court or appellate court that would guarantee no bad laws.*
4. *The cases that may make bad law are intermediately difficult.*
5. *Bad laws are Pareto-dominated.*

Proofs

Proof of Remark 2. Note that $x < -\epsilon \implies x_t < 0$ and $x > \epsilon \implies x_t > 0$, so setting $r = 0$ would guarantee the realization of the disposition component of HC's payoff function. Because $r = 0$ is the unique maximizer of the rule component of HC's payoff function, it follows that $r = 0$ is HC's uniquely optimal action whenever $x \notin [-\epsilon, \epsilon]$. □

All statements in Proposition 4 follow from the following characterization of perfect Bayesian equilibrium:

Proposition 5. *The following profile of strategies and beliefs characterizes a perfect Bayesian equilibrium of the judicial-hierarchy game with perfectly inclusive doctrine.*

1. *If $x \geq \epsilon$ or $x < -\epsilon$ then HC sets $r = 0$ and LC sets $x' = x$.*

2. *If* $x \in [-\epsilon, 0)$ *then HC sets* $r = 0$ *and LC sets* $\begin{cases} x' = x & \textit{if } x_t < L \\ x' \geq 0 & \textit{if } x_t \geq L \end{cases}$.

3. *Define* $x^* = \begin{cases} x_m^* & \textit{for } L \geq x_m^* + \epsilon \\ \dfrac{2\epsilon + L + e_h - \sqrt{L^2 + e_h^2 + 6Le_h}}{2} & \textit{for } L < x_m^* + \epsilon \end{cases}$ *and* $x_m^* = \dfrac{\epsilon^2}{e_h + \epsilon}$.

4. *If* $x \in [0, x^*]$, *then*

   - *HC sets* $r = \begin{cases} 0 & \textit{if } EU_{HC}(r = 0) \geq EU_{HC}(r = x') \\ x' & \textit{if } EU_{HC}(r = 0) < EU_{HC}(r = x') \end{cases}$.
     In particular, on the path of play HC sets $r = 0$.
   - *LC sets* $x' = \begin{cases} x - \epsilon & \textit{if } x_t < L \\ x & \textit{if } x_t \geq L \end{cases}$.

5. *If* $x \in (x^*, \epsilon)$ *then*

   - *HC sets* $\begin{cases} r = 0 & \textit{if} \quad EU_{HC}(r = 0) > EU_{HC}(r = x') \\ r = x' & \textit{if} \quad EU_{HC}(r = 0) < EU_{HC}(r = x') \\ r = \begin{cases} 0 \textit{ w/ prob.} & p \\ x' \textit{ w/ prob.} & 1 - p \end{cases} & \textit{if} \quad EU_{HC}(r = 0) = EU_{HC}(r = x') \end{cases}$
     where $p = \dfrac{\epsilon - x}{e_\ell + \epsilon - x}$.
     In particular, on the path of play HC sets $r = 0$ *if* $x' = x$ *and*
     $r = \begin{cases} 0 & \textit{w/ prob.} \quad p \\ x - \epsilon & \textit{w/ prob.} \quad 1 - p \end{cases}$ *if* $x' = x - \epsilon$.
   - *LC sets* $\begin{cases} x' = x & \textit{if} \quad x_t \geq L \\ x' = \begin{cases} x - \epsilon & \textit{w/ prob.} \quad \pi_1 \\ x & \textit{w/ prob.} \quad 1 - \pi_1 \end{cases} & \textit{if} \quad x_t < 0 \\ x' = \begin{cases} x - \epsilon & \textit{w/ prob.} \quad \pi_2 \\ x & \textit{w/ prob.} \quad 1 - \pi_2 \end{cases} & \textit{if} \quad x_t \in [0, L) \end{cases}$

$$\text{where } \pi_1 = \begin{cases} \pi_2 \dfrac{(\epsilon + x)(x - \epsilon + e_h)}{(\epsilon - x)(\epsilon - x + e_h)} & \text{if } x < L - \epsilon \\ \pi_2 \dfrac{L(x - \epsilon + e_h)}{(\epsilon - x)(\epsilon - x + e_h)} & \text{if } x \geq L - \epsilon \end{cases}.$$

6. *HC's beliefs on path are given by Bayes's rule.*

   Off path, $\Pr(x_t < 0 | x' \geq 0) = 0$ *and if* $x' < 0$ *then* $x_t \sim U[x - \epsilon, \min\{L, x + \epsilon\})$.

Proof. The proof verifies that the players' strategies are best responses for different realizations of x. To begin, note that for any x and any x', all rule choices except $r = 0$ and $r = x'$ are dominated for HC. If $x' \geq 0$ then all $r > x'$ are strictly worse for HC than $r = x'$ because $r = x'$ yields the same disposition and a higher rule utility, and all $r < x'$ are strictly worse than $r = 0$ because $r = 0$ yields the same disposition and a higher rule utility. If $x' < 0$ then all $r < x'$ are strictly worse than $r = x'$ because $r = x'$ yields the same disposition and a higher rule utility, and all $r > x'$ are strictly worse than $r = 0$ because $r = 0$ yields the same disposition and a higher rule utility. Therefore, when checking HC's deviations, it is sufficient to check $r = 0$ and $r = x'$.

First consider the case $x \geq \epsilon$ or $x < -\epsilon$. If $x \geq \epsilon$ then $x_t \geq 0$ and $x' \geq 0$; and if $x < -\epsilon$ then $x_t < 0$ and $x' < 0$, so in both cases HC's uniquely optimal action is $r = 0$, regardless of what LC does. Therefore, $x' = x$ is a best response for LC.

Next consider the case $x \in [-\epsilon, 0)$. If $x \in [-\epsilon, 0)$ and $r = 0$ and $x_t < L$ then $U_{LC}(x' = x) = -L + e_\ell = U_{LC}(x' < 0) > -L = U_{LC}(x' \geq 0)$. If $x \in [-\epsilon, 0)$ and $r = 0$ and $x_t \geq L$ then $U_{LC}(x' \geq 0) = -L + e_\ell > -L = U_{LC}(x' < 0)$. So LC has no profitable deviation. As for HC, if $x \in [-\epsilon, 0)$ and $x' = x$ then $EU_{HC}(r = 0) = e_h \Pr(x_t < 0)$ and $EU_{HC}(r = x') = x + \Pr(x_t \geq 0)$. Consider two cases separately: (1) $L \geq x + \epsilon$, (2) $L < x + \epsilon$. In the first case $EU_{HC}(r = 0) = \left(\dfrac{\epsilon - x}{2\epsilon}\right) e_h > x + \left(\dfrac{\epsilon + x}{2\epsilon}\right) e_h = EU_{HC}(r = x')$. In the second case $EU_{HC}(r = 0) = \left(\dfrac{\epsilon - x}{L - x + \epsilon}\right) e_h > x + \left(\dfrac{L}{L - x + \epsilon}\right) e_h = EU_{HC}(r = x')$. If $x \in [-\epsilon, 0)$ and $x' \geq 0$ on the path of play then $EU_{HC}(r = 0) = e_h = \max U_{HC}$. Off the path of play, if $x' < 0$ and $x' \neq x$ then $r = 0$ is a best response by the same reasoning as if $x' = x$; and if $x' \geq 0$ then $EU_{HC}(r = 0) = e_h = \max U_{HC}$. We conclude that the strategies are best responses when $x \in [-\epsilon, 0)$.

Next consider the case $x \in [0, \epsilon)$. Define x^* as the value of x such that if LC sets

$x' = \begin{cases} x - \epsilon & \text{if } \; x_t < L \\ x & \text{if } \; x_t \geq L \end{cases}$, a strategy which will be denoted σ_L, then $EU_{HC}(r = x'|\sigma_L, x' = x - \epsilon, x = x^*) = EU_{HC}(r = 0|\sigma_L, x' = x - \epsilon, x = x^*)$. It will help to consider three cases separately: (1) $L > 2\epsilon$, (2) $L \in (x_m^* + \epsilon, 2\epsilon]$ (where x_m^* will be defined as the value of x^* for certain values of L), (3) $L \in (0, x_m^* + \epsilon]$.

In the first case $x \in [0, \epsilon) \implies x_t < L$, so σ_L becomes $x' = x - \epsilon$. Then setting $EU_{HC}(r = 0) = EU_{HC}(r = x')$ and solving for x yields $x^* = \dfrac{\epsilon^2}{\epsilon + e_h} \equiv x_m^*$. First consider $x \in [0, x_m^*]$. HC's strategy of choosing between $r = 0$ and $r = x'$ is optimal, as argued at the beginning of the proof. In particular, on the path of play, it follows from the definition of x_m^* that $EU_{HC}(r = 0) \geq EU_{HC}(r = x - \epsilon)$. As for LC, $U_{LC}(x' = x - \epsilon) = -L + e_\ell$. Now if LC sets $x' \geq 0$ then, given off-path beliefs, HC would set $r = 0$, so $U_{LC}(x' \geq 0) = -L$. And if LC sets $x' \in (x - \epsilon, 0)$ then $U_{LC} = \begin{cases} -L + e_\ell & \text{if } \; r = 0 \\ x' - L & \text{if } \; r = x' \end{cases}$. So LC does not have a profitable deviation, and the parties' strategies are best responses for $x \in [0, x_m^*]$. If $x \in (x_m^*, \epsilon)$ then for LC to randomize between $x' = x$ and $x' = x - \epsilon$ we must have $EU_{LC}(x' = x) = EU_{LC}(x' = x - \epsilon)$, which is to say $-L = \Pr(r = 0|x' = x - \epsilon)(-L + e_\ell) + \Pr(r = x - \epsilon|x' = x - \epsilon)(x - \epsilon - L)$, which, denoting $\Pr(r = 0|x' = x - \epsilon)$ by p, yields $p = \dfrac{\epsilon - x}{\epsilon - x + e_\ell}$. It remains to verify that LC does not have a profitable deviation from $x' = x$ and $x' = x - \epsilon$. If LC sets $x' \geq 0, x' \neq x$ then, given off-path beliefs, HC would set $r = 0$, so $U_{LC}(x' = x) = U_{LC}(x' \geq 0, x' \neq x)$. If LC sets $x' \in (x - \epsilon, 0)$ then, given off-path beliefs and by the definition of x_m^*, HC would set $r = x'$, so $U_{LC}(x' \in (x - \epsilon, 0)) = x' - L < -L = U_{LC}(x' = x)$. So LC does not have a profitable deviation. For HC to randomize between $r = 0$ and $r = x'$ in response to $x' = x - \epsilon$ we must have $EU_{HC}(r = 0) = EU_{HC}(r = x - \epsilon)$, which is to say $e_h \Pr(x_t < 0) = (x - \epsilon) \Pr(x_t < 0) + (x - \epsilon + e_h) \Pr(x_t \geq 0)$, which yields $\Pr(x_t < 0|x' = x - \epsilon) = \dfrac{x - \epsilon + e_h}{2e_h}$. Then, using Bayes's rule, and defining $\pi_1 \equiv \Pr(x' = x - \epsilon|x_t < 0)$ and $\pi_2 \equiv \Pr(x' = x - \epsilon|x_t \geq 0)$, we obtain $\pi_1 = \dfrac{(x + \epsilon)(x - \epsilon + e_h)}{(\epsilon - x)(e_h + \epsilon - x)}\pi_2$. (It can be verified that $\pi_1 > \pi_2 \; \forall x > x_m^*$.) It follows that HC's randomizing between $r = 0$ and $r = x - \epsilon$ is a best response to $x' = x - \epsilon$. To see that $r = 0$ is a best response to $x' = x$, note that $EU_{HC}(r = 0|x' = x) = e_h \Pr(x_t \geq 0|x' = x) = e_h \dfrac{(1 - \pi_2)(\epsilon + x)}{(1 - \pi_2)(\epsilon + x) + (1 - \pi_1)(\epsilon - x)}$

and $EU_{HC}(r = x|x' = x) = -x + e_h \Pr(x_t < 0|x' = x) = -x + e_h(1 - \Pr(x_t \geq 0|x' = x))$. Because $\pi_2 < \pi_1$ and $\epsilon + x > \epsilon - x$, it follows that $\Pr(x_t \geq 0|x' = x) > 1/2$, so $EU_{HC}(r = 0|x' = x) > EU_{HC}(r = x|x' = x)$. We see that the strategies are best responses in the first case.

Now consider the second case, $L \in (x^*_m + \epsilon, 2\epsilon]$. If $x \leq x^*_m$ then $x_t < L$, so σ_L becomes $x' = x - \epsilon$ and, by the same reasoning as in the last paragraph, the players' strategies are best responses. For $x > x^*_m$ we check LC and HC's strategies in turn. If $x > x^*_m$ and $x_t < L$ then for LC to randomize between $x' = x$ and $x' = x - \epsilon$ we must have $EU_{LC}(x' = x) = EU_{LC}(x' = x - \epsilon)$, which by the same reasoning as above yields $p = \dfrac{\epsilon - x}{\epsilon - x + e_\ell}$. Moreover, by the same reasoning as above, LC does not have a profitable deviation from $x' = x$ or $x' = x - \epsilon$. If $x > x^*_m$ and $x_t \geq L$ then $U_{LC}(x' = x) = -L + e_\ell$. If LC sets $x' \geq 0, x' \neq x$ then, given off-path beliefs, HC would respond by $r = 0$, so $U_{LC}(x' \geq 0, x' \neq x) = -L + e_\ell = U_{LC}(x' = x)$. If LC sets $x' \in (x - \epsilon, 0)$ then, given off-path beliefs, HC would respond by $r = x'$, so $U_{LC}(x \in (x - \epsilon, 0)) = x' - L + e_\ell < U_{LC}(x' = x)$. Lastly, $EU_{LC}(x' = x - \epsilon) = -L + (1 - p)(x - \epsilon + e_\ell) < U_{LC}(x' = x)$. So LC has no profitable deviations. For HC to mix between $r = 0$ and $r = x - \epsilon$ in response to $x' = x - \epsilon$ we must have $EU_{HC}(r = 0) = EU_{HC}(r = x - \epsilon)$, which by the same reasoning as above yields $\Pr(x_t < 0|x' = x - \epsilon) = \dfrac{x - \epsilon + e_h}{2e_h}$. Now we compute the probability of LC's factfinding by Bayes's rule. For $x < L - \epsilon$, the calculation is the same as in the preceding paragraph and yields the same π_1 and π_2. For $x \geq L - \epsilon$, using Bayes's rule and using π'_1 and π'_2 in place of π_1 and π_2 we obtain $\pi'_1 = \dfrac{L(x - \epsilon + e_h)}{(\epsilon - x)(e_h + \epsilon - x)}\pi'_2$. (It can be verified that $\pi'_1 > \pi'_2 \; \forall x > x^*_m$.) Finally we verify that $r = 0$ is HC's best response to $x' = x$. $EU_{HC}(r = 0|x' = x) = e_h \Pr(x_t \geq 0|x' = x)$ and $EU_{HC}(r = x|x' = x) = -x + e_h(1 - \Pr(x_t \geq 0|x' = x))$. Now if $x < L - \epsilon$ then $\Pr(x_t \geq 0|x' = x) = \dfrac{(1 - \pi_2)(\epsilon + x)}{(1 - \pi_2)(\epsilon + x) + (1 - \pi_1)(\epsilon - x)} > 1/2$, so $EU_{HC}(r = 0) > EU_{HC}(r = x')$. If $x \geq L - \epsilon$ then $\Pr(x_t \geq 0|x' = x) = \dfrac{(1 - \pi'_2)L + \epsilon - L + x}{(1 - \pi'_2)L + \epsilon - L + x + (1 - \pi'_1)(\epsilon - x)} > 1/2$, so $EU_{HC}(r = 0) > EU_{HC}(r = x')$. We see that HC has no profitable deviations. The players' strategies are best responses in the second case.

Lastly consider the case $L \in (0, x^*_m + \epsilon]$. First we calculate x^* which, recall, is the value of x that solves $EU_{HC}(r = 0|\sigma_L, x' = x - \epsilon) = EU_{HC}(r = x'|\sigma_L, x' = x - \epsilon)$. The equation yields $e_h \Pr(x_t < 0|x' = x - \epsilon) = x - \epsilon + e_h \Pr(x_t \geq 0|x' = x - \epsilon)$. We now use Bayes's rule to compute $\Pr(x_t \geq 0 | x' = x - \epsilon)$. Noting that $x < L - \epsilon \implies x < x^*_m \implies EU_{HC}(r = 0) > EU_{HC}(r = x - \epsilon)$, we need only consider the case where $x \geq L - \epsilon$, which yields $\Pr(x_t < 0 | x' = x - \epsilon) = \dfrac{\epsilon - x}{\epsilon - x + L}$. Plugging this back into the equation for x^* yields $x^* = \dfrac{2\epsilon + L + e_h - \sqrt{L^2 + e_h^2 + 6Le_h}}{2}$. Next we verify that the players' strategies are best responses. If $x \leq x^*$ then the strategies are best responses by the same reasoning as before. For $x > x^*$ we check LC and HC's strategies in turn. For LC, the derivation of p and the showing that LC does not have a profitable deviation are the same as in the preceding paragraph. For HC, we consider $x' = x - \epsilon$ and $x' = x$ in turn. To randomize between $r = 0$ and $r = x - \epsilon$ in response to $x' = x - \epsilon$ requires $EU_{HC}(r = 0 | x' = x - \epsilon) = EU_{HC}(r = x - \epsilon | x' = x - \epsilon)$. Using Bayes's rule as before, noting that $x > L - \epsilon$, and denoting $\pi_1' = \Pr(x' = x - \epsilon | x_t < 0)$ and $\pi_2' = \Pr(x' = x - \epsilon | x_t \in [0, L))$, we obtain $\pi_1' = \dfrac{L(x - \epsilon + e_h)}{(\epsilon - x)(e_h + \epsilon - x)}\pi_2'$. Note that this is the same as π_1' and π_2' calculated in the preceding paragraph. Now if $x' = x$ then we must show that $EU_{HC}(r = 0) \geq EU_{HC}(r = x)$. Note that $EU_{HC}(r = x) = \left(\dfrac{(1 - \pi_1')(\epsilon - x)}{(1 - \pi_1')(\epsilon - x) + \epsilon + x - \pi_2' L}\right) e_h - x$ and $EU_{HC}(r = 0) = \left(1 - \dfrac{(1 - \pi_1')(\epsilon - x)}{(1 - \pi_1')(\epsilon - x) + \epsilon + x - \pi_2' L}\right) e_h$. Noting that $\dfrac{(1 - \pi_1')(\epsilon - x)}{(1 - \pi_1')(\epsilon - x) + \epsilon + x - \pi_2' L} < 1/2$, it follows that $EU_{HC}(r = 0) > EU_{HC}(r = x)$. This concludes the proof of Proposition 5. □

Proof of Proposition 4. The statements follow immediately from the equilibrium characterization in Proposition 5:

1. Cases in the range $x \in (x^*, \epsilon)$ make bad law ($r \neq 0$) with positive probability.
2. Cases where $x \notin [-\epsilon, \epsilon]$ never make bad law. (More generally than in the specified equilibrium, this statement follows from Remark 2.)
3. Cases in the range $x \in (x^*, \epsilon)$ make bad law with positive probability whenever $e_h \neq 0$ and $e_\ell \neq 0$.
4. The region where bad law is made with positive probability is an intermediate region of the public signal and hence of case facts, namely $x \in (x^*, \epsilon)$.
5. When bad law is made, the result is $r = x - \epsilon < 0$, which is Pareto-dominated by some $r \in [0, L]$, for example by $r = 0$. □