Precedential Cascades: A Critical Appraisal

Eric Talley


Sponsored by the John M. Olin Foundation

University of Southern California Law School
Los Angeles, CA 90089-0071

This paper can be downloaded without charge from the Social Science Research Network electronic library at http://papers.ssrn.com/paper.taf?abstract_id=160277
PRECEDENTIAL CASCADES:
A CRITICAL APPRAISAL

Eric Talley†
April 1999

CONTENTS

INTRODUCTION .............................................................. 1

I. ARE PRECEDENTIAL CASCADES POSSIBLE? ............................. 7
   (a) A Basic Framework ............................................ 8
   (b) The Emergence of a Precedential Cascade ..................... 13
   (c) Significance .................................................... 16

II. ARE PRECEDENTIAL CASCADES PLAUSIBLE? .......................... 19
   (a) Rule-boundedness. .............................................. 20
   (b) Decisional Opacity. .......................................... 21
   (c) Judicial Homogeneity ....................................... 24
   (d) Short Judicial Tenures. ..................................... 26
   (e) Flat Hierarchies. ............................................. 27
   (f) Population Stationarity. .................................... 28

III. ARE PRECEDENTIAL CASCADES VERIFIABLE? ........................ 29
   (a) Empirical Evidence .......................................... 30
   (b) Anecdotal Evidence ......................................... 34

CONCLUSION ................................................................. 37

APPENDIX .......................................................................... 39

†Associate Professor, University of Southern California Law School, University Park, Los Angeles, CA 90089-0071. Email: etalley@law.usc.edu. © Eric L. Talley, 1999. This article grew out of an original working paper entitled, On Common Law Learning, Rational Emulation, and the Evolution of Legal Rules (USC mimeo, January, 1996). Thanks to Scott Altman, Ian Ayres, Antonio Bernardo, Erwin Chemerinsky, Marshall Cohen, David Cruz, Mark Grady, Ariela Gross, Gillian Lester, Susanne Lohmann, Tom Lyon, Paul Mahoney, Ed McCaffery, Nolan McCarty, Michael Shapiro, Jeff Strnad and Ivo Welch for helpful comments and discussions on the earlier version. Thanks as well to Cindy Janis and Edgar Park for excellent research assistance. All errors are mine.
PRECEDENTIAL CASCADES:
A CRITICAL APPRAISAL

Eric Talley

ABSTRACT

A number of legal scholars have recently posited a “group herding” account of legal institutions and precedent. Their arguments draw heavily from the emergent literature on information cascades—a set of theories that attempts to explain certain pathologies of group behavior. Explicitly, cascade theory demonstrates how rational individuals, when making ostensibly individual decisions, may ignore their own private inclinations, choosing to emulate others instead. When aggregated, such behavior can lead to significant errors of judgment on behalf of the entire group, particularly if exacerbated by other cognitive effects. If legal precedent is a species of cascade, then the consequences for legal theory are profound. Indeed, not only would such a theory call into question the wisdom of celebrated “watershed” cases from the last century, but it would also severely compromise virtually any positive account of law that conceives of precedent as a central mechanism for judicial learning.

This article critically analyzes the viability of a cascade theory of precedent, asking whether such phenomena are (a) possible, (b) plausible, and (c) empirically verifiable. Although I find that a precedential cascade is certainly possible as a matter of theory, the necessary conditions for its occurrence are either extremely rare or easily avoidable. Moreover, even the more modest task of diagnosing when precedential cascades occur may be unavailing: for the outward symptoms of a cascade are essentially indistinguishable from those of more plausible (and less troubling) behavioral hypotheses.
INTRODUCTION

About a half century ago, social scientists began to develop what was to become the intellectual heart of neoclassical economics. Under the broad rubric of "general equilibrium theory," these scholars formalized, in varying degrees of mathematical sophistication, the longstanding intuition that competitive markets could convert apparent disarray and fragmentation into order and harmony. More explicitly, general equilibrium theory demonstrated how a decentralized collection of self-interested individuals could, through competitive market transactions, allocate scarce goods and services in a socially-efficient manner. An equally powerful corollary attended this central insight: that the prices that emerge in such markets convey valuable public information about resource scarcity to individuals possessing little more than "dispersed bits of incomplete and frequently contradictory knowledge." By many accounts, these dual characteristics of efficiency and information transmission constitute a legacy that is central to much of modern economic thought.

In light of this intellectual legacy, it is hardly surprising that early law-and-economics scholars entertained the possibility that the common law could function in a manner similar to competitive markets, effecting order from chaos. The specifics

---


2 This proposition has come to be known as the First Theorem of Welfare Economics. It is closely related (perhaps unsurprisingly) to the Second Theorem of Welfare Economics, which states (under certain conditions on preferences) that any Pareto efficient allocation can be supported by a competitive equilibrium with appropriate lump sum transfers. See Hal Varian, MICROECONOMIC ANALYSIS 200-25 (1984).

3 F.A. Hayek, The Use of Knowledge in Society, 35 AMER. ECON. REV. 519 (1945).

4 Even in so-called "New Keynesian" economics, a number of theories abound in which price is seen as a critical transmission device for information. See, e.g., Joseph Stiglitz, The Causes and Consequences of the Dependence of Quality on Price, 25 J. ECON. LIT. 1, 2 (1987) (arguing that price may serve an important signal of quality).
of their story would take a slightly different form, of course. Judges and litigants would supplant firms and consumers as the central economic actors; precedents and rights would replace prices and quantities as the focal equilibrium outcomes. But the underlying argument was otherwise similar—positing first that common-law precedent tends over time to converge to rules which are economically efficient, and second that it does so in a decidedly non-orchestrated fashion—obviating the up-front costs that plague statutory promulgation in favor of a less costly, learning-by-doing protocol.

On first glance, the market analogy has a number of appealing characteristics. Indeed, judges share at least some attributes with actors in a decentralized economy. They possess limited information about the world; they pursue ends which need not coincide with broader social objectives (however defined); and they often pursue their respective ends independently rather than collectively. Moreover, just as prices affect individual consumption, production, and investment decisions, so too can legal rules constrain and shape individual behavior. Buttressing these similarities, early advocates of the market analogy (most notably Richard Posner) identified some striking examples within the common law in which we observe rules that seem (at least roughly) to correspond with efficiency concerns.

Nevertheless, among contemporary legal scholars, the analogy between market behavior and legal evolution remains relatively tenuous. Over the last twenty years, many detractors of the efficiency hypothesis have challenged its foundational premises. Some, for example, have noted that judges (often by design) are more insulated from the inefficient consequences of their decisions than are analogous

---


6 Cite.
market participants. This added protection, in turn, affords them the opportunity to pursue ends that need not be related systematically to efficiency. Others have registered even stronger criticisms, noting that even if courts were pre-disposed towards efficiency concerns, there is no guarantee that those who actually litigate important matters will constitute an unbiased sample from the relevant population. Operating from such a skewed sample, the argument goes, judges may stand little chance of arriving at an efficient rule, even if they wanted to.

Quite recently, however, a number of scholars have developed a more novel account of precedential evolution—one that poses an even greater challenge to the efficiency thesis: the theory of rational herding, often called “information cascades.” Originating from simple models of learning, cascade theory lends some interesting and provocative insights about the relationship between individual rationality and group behavior. In particular, it offers an account of why seemingly rational individuals, when making (ostensibly) independent decisions, might ignore their own inclinations, preferring instead to emulate others. More specifically, the cascades literature posits that strategic actors may rationally prefer emulation, presuming (perhaps incorrectly) that their own information is inadequate as measured against the stock of that

---


In recent years, a number of different models of "rational herding" have emerged in the finance and organizational economics literature. See, e.g., Bikhchandani, Hirshleifer & Welch, A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades, 100 J. Polit. Econ. 992 (1992); Banerjee, A., A Simple Model of Herd Behavior, 107 Q. J. Econ. 797 (1992); Zwiebel, J., Corporate Conservatism and Relative Compensation, 103 J. Political Econ. 1 (1995); Scharfstein, D. & Stein, J, Herd Behavior and Investment, 80 Amer. Econ. Rev. 465 (1990).

Moreover, once an information cascade begins, other forces (such as reputational concerns, pressures to conform, and social "entrepreneurs" eager to exploit group pathologies) may amplify its longevity and severity. In the aggregate, then, serial decisions to follow the herd may lead to a form of group stagnation, curtailing the learning process not only for the marginal actor, but also for all who succeed her. Groups of actors caught within a cascade are therefore prone to severe lapses in judgment, and one can never be confident that the resulting trajectory of behavior—no matter how stable—reflects a desirable social policy.

Although the first applications of the herding literature were to extra-legal phenomena (such as popular fads, political revolutions and stock market crashes), its relevance for the theory of common-law evolution are apparent. Indeed, on at least a descriptive level, legal precedents have the makings of information cascades. Numerous legal scholars (beginning as early as Oliver Wendell Holmes) have maintained that the formation of precedent is symptomatic of judicial learning: once judges gain sufficient familiarity with a subject matter (the argument goes)—transmitted through successive holdings in "similarly situated" cases—a binding rule may crystalize for all future cases, even if the particularized facts of those cases suggest a different result. In this sense, just as in an information cascade, a series of like-minded holdings may transmit information from preceding to succeeding judges—information that eventually obviates the need for further inquiry.

10In recent years, a number of different models of "rational herding" have emerged in the finance and organizational economics literature. See, e.g., Bikhchandani, Hirshleifer & Welch, A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades, 100 J. Polit. Econ. 992 (1992); Banerjee, A., A Simple Model of Herd Behavior, 107 Q. J. Econ. 797 (1992); Zwiebel, J., Corporate Conservatism and Relative Compensation, 103 J. Political Econ. 1 (1995); Scharfstein, D. & Stein, J, Herd Behavior and Investment, 80 Amer. Econ. Rev. 465 (1990).


12Oliver Wendell Holmes's famous lectures on criminal and tort law probably sowed the seeds for what was to become this "learning" account of the common law within the law and economics literature. See Oliver W. Holmes, Jr., The Common Law 39-129 (1881). Though
If common law precedent is in fact a type of information cascade, it would represent the strongest refutation yet of the common law efficiency hypothesis. Indeed, it would suggest that even if judges are predisposed towards efficiency, and even if they do not face a biased selection of cases, precedents might still frequently diverge from the most efficient legal rule. Moreover, a theory of precedential herding would force us to rethink the coherence of virtually any jurisprudential theory of precedent that conceives of the common law as a mechanism for judicial learning—be it economic or otherwise. If precedent represents a weak or impoverished learning device, then a common law system of adjudication seems unlikely to produce reliable results. More concretely, however, this account of precedent (if true) would shed considerable light on why some legal doctrines are created (or destroyed) at an extremely rapid pace. It might explain, for example, why solitary decisions such as 

Lochner v. New York\textsuperscript{13} succeeded in sowing the seeds for a string of succeeding like-minded (and non-self-conscious) opinions, only to meet its demise when its tenuous foundational underpinnings are finally exposed. On the other hand, a herding theory of precedent might cause us to question the wisdom of surviving doctrines that emanate from watershed cases such as Brown v. Board of Education,\textsuperscript{14} as we could

Holmes did not cast his account within the rhetoric of economic efficiency, there is a fairly close parallel. Id. at 1-2 (“The substance of the law at any given time pretty nearly corresponds, so far as it goes, with what is then understood to be convenient. . . .”). The Holmesian account is reflected in economic terms by Richard Posner, who writes:

A system of decision according to precedent has [an] economizing property: It [enables] the parties to a case, and the tribunal also, to use information that has been gathered (often at considerable expense) in previous cases. If it has been held in 20 cases that an electric crossing signal is a required (cost-justified) precaution at busy railroad crossings, the marginal gain in knowledge of the relevant costs and values from incurring the expense of a trial in the twenty-first case may be smaller than the expense. A rule of the common law emerges when its factual premises have been so validated by repeated testing in litigation the traditional expenditures on proof and argument would exceed the value of the additional knowledge, and hence the less valuable will be adherence to precedent. The authority and information aspects of decision making according to precedent thus converge in suggesting that the practice will be consistently followed only in highly stable societies.


\textsuperscript{13}198 US 45 (1905)\n
\textsuperscript{14}347 US 483 (1954).
no longer be sure that their apparent durability is a result of sound foundational premises or historical accidents.

In this article, I endeavor to evaluate whether the emergent theory of cascades represents a cogent positive account of legal evolution. My conclusions fall far short of an endorsement: While it is certainly possible for precedent to manifest some cascade-like characteristics, the necessary conditions for such phenomena to occur are either unlikely or easily avoidable (or both) within most practical contexts. In particular, many existing institutional practices (such as judicial tenures, written opinions, and the hierarchical appeals processes) tend to reduce the likelihood of a “bad” precedential cascade. At the same time, however, courts and legal scholars would be well-advised to remain aware the potential dangers of herding, as the populations that legal rules endeavor to regulate may themselves be prone to problematic herding behavior.

I develop my argument in three parts. Part I asks whether it is theoretically possible for judicial precedent to constitute an information cascade. My answer to this question is “yes,” and I demonstrate how by analyzing a stylized model of adjudication, in which a sequence of judges rules on a sequence of similarly-situated cases. The example helps to demonstrate how a “precedential cascade” might emerge if at any point in time the relevant case history becomes sufficiently monolithic. Moreover, I demonstrate how the resulting precedent need not coincide, ironically, with the underlying normative goals of the judges creating it, thereby exposing the principal theoretical danger of such cascades. In Part II, I turn to the more practical question of whether such herding behavior is a plausible positive account of law. Here my answer is somewhat more critical. Explicitly, I consider six necessary preconditions for precedential cascades to occur: Rule-boundedness, decisional opacity, judicial homogeneity, flat hierarchies, short judicial tenure, and population stationarity. I argue that not only is each of these preconditions far from ubiquitous, but that their confluence is likely to be especially rare. Moreover, a number of them are easily avoidable with a prudently designed judicial system, the features of which are already largely present within the current institutional regime. Finally, Part III turns to empirical concerns, and explores how one might go about verifying whether
legal cascades exist in at least some circumstances. Here my conclusions here are perhaps the most pessimistic. For even if precedential cascades seemed plausible on a priori grounds, it is difficult (and likely impossible) test for their existence against any number of plausible alternative hypotheses. A technical appendix follows the conclusion of this article for those interested in the formal development of the intuitions presented in the text.

I. ARE PRECEDENTIAL CASCADES POSSIBLE?

In order to conduct a critical analysis of the cascade theory of precedent, it is first necessary to understanding of how herding devices operate. This Part endeavors to do just that, demonstrating what precisely an informational cascade is, and how one might occur within a jurisprudential environment. The discussion below analyzes a simple numerical model of adjudication, in which a series of like-minded judges act sequentially in ruling on a series of similarly-situated cases. The judges are presumed rational, but incompletely informed about the world, and thus may use the holdings of their predecessors in action to inform their own beliefs. An analysis of this example yields two interesting results: First, as judges learn information from previous holdings, they may rationally begin to treat such holdings as binding on them, even if not formally required to do so, and even if the case they actually hear suggests a contrary outcome. Second, the incentive to emulate one another may stagnate the learning process, creating a precedent that fails to coincide with the normative objectives of each judge.

Before proceeding with the model, a few caveats are in order. First, the analysis that follows presumes throughout that there is no external “rule” of binding precedent which compels judges to decide cases in a manner identical to their predecessors in action. Indeed, cascade theory is a positive account of group behavior, and thus it is inappropriate to assume (for various reasons emanating from

---

15In what follows, I shall presume that this common goal is efficiency-oriented in nature. However, a similar analysis would flow from any normative goal that depends, at least in part, on the information that one might learn from each successive case.
outside the model) that such precedent is binding. Rather, the ultimate question of interest here is under what circumstances will judges rationally choose to emulate predecessors, even if the case before them suggests an inconsistent result. To be sure, there are many other positive theories of precedent. But they are immaterial for purposes of understanding the account of precedent as a cascade.

Second, as with any model, the analysis below is a simplified description of the world, designed to expose the fundamental reasons why an information cascade might emerge. It is therefore important to distinguish between the model’s simplifying assumptions (i.e., those that make the analysis more tractable without altering any qualitative results) and critical assumptions (i.e., those that do both). Indeed, a subsequent Part of this article examines six critical assumptions of herding models, and argues that their confluence in practice may be extremely implausible.16

(a) A Basic Framework

Consider a single jurisdiction with a single court in which legal rules evolve from generation to succeeding generation. In each generation (or period), the jurisdiction appoints a presiding judge who is a sophisticated, rational Bayesian.17

16See Part II, infra.

17The descriptive term “Bayesian” asserts roughly that each judge uses available information to update and refine her beliefs about the world. In the context of this article, judges use past and current cases to hone their beliefs about the composition of the population they must regulate through choice of a legal rule.

The process of updating follows what is known as “Bayes rule.” The rule states that the probability of an event \( A \) occurring, conditional on knowing that some other event \( B \) has occurred (or in short hand, \( \Pr(A|B) \)) can be derived from a combination of the respective “base rate” probabilities of \( A \) and \( B \) (or \( \Pr(A) \) and \( \Pr(B) \)) and the “reverse conditional” probability that \( B \) occurs, conditional on knowing that \( A \) has occurred (or \( \Pr(B|A) \)). Explicitly, these four probabilities are related to one another according to the following formula:

\[
\Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)}
\]

Thus, for example, suppose that one were attempting to use the outcome of a diagnostic test to infer whether an individual was carrying a deadly virus. Suppose further that when administered to people who are known to have the virus, the test yields a positive result 50% of the time. Moreover, among the general population, the test yields a positive result 25% of the time, and 10% of the general population carries the virus. Using Bayes rule, it is possible to calculate the probability that
motivated solely by economic efficiency concerns. The tenure of each judge lasts exactly one period, at which time she retires and is replaced by a successor, who is similarly sophisticated, rational, Bayesian, and efficiency-minded. Denote the series of judges as time passes by \( J_1, J_2, J_3, \ldots \) and so on.

The principal task for each succeeding judge (or \( J_i \)) is to announce a legal rule that shall govern a given “class” of cases in the jurisdiction during the coming period.\(^{18}\) To make things a bit more concrete, I will borrow a classic “least cost risk avoider” account from the law-and-economics literature on accidents (though the intuition readily can be recast in other areas as well). Suppose that each succeeding judge must announce a rule to govern some potentially injurious activity, whose possible harm to victims is high, say $1 million. It is not particularly important what the activity is, other than to allow for either the potential injurer or the potential victim to engage in harm-avoiding investments (specified at greater length below).\(^{19}\) The judge must decide between two potential rules: Under the first (the “pro-victim” rule), potential injurers are strictly liable for injuries that occur as the result of an accident.\(^{20}\) Under the second (the “pro-injurer” rule), the risk of accident falls wholly on potential victims of the harmful activity. (For shorthand reference in what follows, I shall at times denoted the rule announced in each period by \( y_i \), such that a decision

\[
\frac{Pr\{\text{Positive Test} | \text{Virus}\}}{Pr\{\text{Positive Test}\}} = \frac{Pr\{\text{Positive Test} | \text{Virus}\} \cdot Pr\{\text{Virus}\}}{Pr\{\text{Positive Test}\}} = \frac{(0.5) \cdot (0.1)}{(0.25)} = 20\%
\]

\(^{18}\)As noted above, this article does not take up at length the question of what makes cases sufficiently similar to place them within the same class (though this may well be an important classificatory question). \textit{See supra} note \_._

\(^{19}\)One example of such an activity is a decision of whether to place a duty of care on elevator manufacturers (who might avoid hazards by installing safety equipment on elevators) or leave potential victims to bear the risk of harm (who might avoid harm by using the stairs instead of elevators).

\(^{20}\)A similar type of analysis is possible if the applicable duty of care rule subjected the injurer to a type of negligence standard. However, most of my analysis applies principally to rule-bound adjudication.
to favor the pro-victim rule corresponds to setting $y_i = 1$, while the pro-injurer rule sets $y_i = 0$). Once the rule is chosen, however, it will apply uniformly to all victim-injurer pairs within the jurisdiction for that period.

Which legal rule applies is important from a social-welfare standpoint, because it identifies the ultimate risk bearer for the activity, who therefore has the largest incentive to engage in precautionary measures to lower the probability of an accident.\textsuperscript{21} To concentrate on this element of the judge's problem, assume that if neither party takes a precaution, then an injury-producing accident is quite likely, at say twenty-five percent. Conversely, should \textit{either} side exercise precaution, the likelihood of harmful accident falls very close to (but slightly above) zero.\textsuperscript{22} Precaution is costly, however, for both potential injurers and potential victims, and as such, the determination of the efficient rule is tantamount to placing the risk on the shoulders of the “least-cost risk avoider.”\textsuperscript{23}

Though this efficiency goal seems rather elementary, each judge's task is complicated by the fact that the relevant population proscribed by the legal rule is \textit{heterogeneous} in nature. Explicitly, suppose that all potential injurers face a cost of precaution of $5,000, while victims can come in one of two flavors: “High-cost” victims, who must bear a relatively high cost of $6,000 to avoid harm; and “Low cost” victims, who face a relatively low avoidance cost of $4,000. In particular, suppose that the fraction of high-cost victims within the jurisdictional population is

\textsuperscript{21}As first noted by Guido Calabresi, the normative efficiency goal of tort law is that the rules on liability should be structured so as to minimize the sum of precaution, accident, and administration costs. \textit{See} Calabresi, \textit{The Cost of Accidents: A Legal and Economic Analysis} (1970); Cooter & Ulen (1988), at 347.

\textsuperscript{22}Thus, some small fraction of victims suffer a harm despite precautions taken by either or both sides, and thus some accidents occur and reach the court system in each period. If precautions reduced the frequency of accidents to zero, then no litigation would occur and there would be no story to tell about common law learning. To keep things simple, I have implicitly assumed here that there are no complementarities in the decision to exercise care. Thus, all the benefits of exercising care are realized if \textit{either} of the parties takes precautions, and there is no added benefit of \textit{both} sides taking precautions. Nothing turns on this assumption.

\textsuperscript{23}See Calabresi, \textit{supra} note 9.
The corresponding fraction of low-cost parties is therefore denoted by \((1-\theta)\), and of course also lies somewhere between 0 and 1.\(^{24}\) Assume (for now) that this proportion \((i.e., \theta)\) remains constant over time.\(^{25}\) The respective victim types, avoidance costs, and population fractions are represented in Table 1 below:

<table>
<thead>
<tr>
<th>Type of Victim</th>
<th>Avoidance Cost</th>
<th>Fraction of Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Cost</td>
<td>$6,000</td>
<td>(\theta)</td>
</tr>
<tr>
<td>Low-Cost</td>
<td>$4,000</td>
<td>((1-\theta))</td>
</tr>
</tbody>
</table>

**Table I. Victim Types, Avoidance Costs, and Population Frequency**

Given the above framework, it is possible to consider the relative efficiency each type of legal rule. On the one hand, it is fairly clear that the pro-victim rule \((i.e., y=1)\) will impose a social cost of $5,000 on each potential injurer. Conversely, the pro-injurer rule \((i.e., y=0)\) will impose a cost of $4,000 for a \((1-\theta)\) proportion of potential victims, and a cost of $6,000 on the remaining \(\theta\) proportion. Hence, it is possible to specify the expected costs that either legal rule imposes on a randomly selected victim-injurer pair, given the relative proportion of victim costs \((i.e., \theta)\):

\[
\text{Exp. Social Costs} = \left\{ \begin{array}{ll}
$4000(1-\theta) + $6000 \theta & \text{under "pro-injurer" rule} \\
$5000 & \text{under "pro-victim" rule}
\end{array} \right.
\]

As noted above, each successive judge \(J_i\) attempts to minimize the expected per-capita social costs during her tenure given her information (which is equivalent to attempting to maximize efficiency\(^{26}\)).

---

\(^{24}\)The corresponding fraction of low-cost parties is therefore denoted by \((1-\theta)\), and of course also lies somewhere between 0 and 1.

\(^{25}\)On relaxing this assumption, see TAN ____-____, infra.

\(^{26}\)To see this, note that the only economic variable in this stylized model that turns on the legal rule is who bears the cost of taking precautions. Under either \(y=0\) or \(y=1\), one of the parties will take precautions. Thus, while the social cost function in the text might also reflect the expected social cost from harm, I exclude it since that cost would be born no matter who bore the cost of precautions.
The heterogeneity of victims complicates matters, because the identity of the least-cost risk avoider depends on whether a potential injurer is paired with a high- or low-cost victim. If the former, then a pro-victim rule is optimal; if the latter, then the opposite holds. Moreover, heterogeneity implies that regardless of the judge’s eventual opinion,\textsuperscript{27} at least some segment of prospective injurer-victim pairs will be “mismatched” with the legal rule. In the face of this uncertainty, then, an efficiency-minded judge will attempt to minimize the costs of mismatches by acquiring enough knowledge (or at least a basis for conjecture) to determine which rule imposes the lowest social cost of risk avoidance on average.\textsuperscript{28}

The judge, therefore, has an incentive (using existing case law and currently-available information) to learn about the characteristics of the injurer-victim population. As is often typical in hard cases, however, no judge has good ex ante information about the true value of $\theta$. Explicitly, assume everyone’s best initial guess about $\theta$ is that it could lie anywhere between 0 and 1 with uniform probability.\textsuperscript{29} It is this lack of knowledge that provides a central reason to hear a case. Indeed, hearing a case affords each successive judge an opportunity to learn more information about the relevant population characteristics, and then to prescribe a rule that best satisfies her jurisdictional preferences during her tenure.\textsuperscript{30}

\textsuperscript{27}The assumption that one rule must apply to all parties is a critical one. In particular, it excludes the possibility of choosing a “tailored” rule in which the victim bears the risk whenever her avoidance costs are $4,000, but the injurer bears the risk whenever the victim’s avoidance costs are $6,000. In some situations, the assumption of a uniform rule is probably warranted. In this case, since the victim and injurer often cannot identify each other ex ante, it is impossible for the injurer to tailor her own behavior in the presence of a “high cost” victim. In other words, since victims are observationally equivalent to injurers, the duty must (for practical reasons) extend uniformly, either to all victims or none.

Nevertheless, there may be a number of situations in which a more tailored rule is available. As I shall argue in Part __, the possibility of a tailored rule may significantly mitigate the dangers of bad information cascades through precedent.

\textsuperscript{28}See Hadfield, supra note __ for a similar approach.

\textsuperscript{29}In other words, I will assume that the prior density function of $\theta$ is uniform on the unit interval.

\textsuperscript{30}Note that if $\theta$ were known with certainty, the optimal legal rule would be $y=0$ if $\theta < 1/2$ and $y=1$ if $\theta > 1/2$. If $\theta = 1/2$, either $y=0$ or $y=1$ would be optimal. In what follows, I will assume a judge who’s beliefs about $\theta$ make her indifferent will simply randomize between the two rules.
In order to capture this intuition,\footnote{See note __, supra.} suppose that at the beginning of her tenure, the judge hears exactly one case, drawn randomly from those in which an accident has occurred.\footnote{Note that the case the judge hears is an unbiased "draw" from the population of potential victims, in which a fraction \((1-\theta)\) are of type 0 while the complementary fraction \(\theta\) are of type 1. In Part II, I consider the effects of allowing judges to hear more than one case. It matters. See TAN ___-___, infra.} In so doing, the judge observes whether the injurer was paired with a low- or high-cost victim. This observation provides the judge with some information about the population, which she may use—along with her knowledge of previous holdings (if any)—to update her belief about the proportion of low- to high-cost victims. She then uses this updated belief to select the rule \((y_i)\) that minimizes prospectively the expected costs of preventing accidents during her tenure.\footnote{Note once again that because the judge is assumed to be motivated solely by efficiency concerns (rather than compensatory goals), the fact that the harm has already occurred makes the sole efficiency value of her opinion prospective in nature. A similar story could be told, however, if the judge were motivated by distributional concerns if she had incomplete information about wealth distributions among injurer-victim pairs.} Once she issues her holding, the judge’s task is complete, and her announced rule governs the jurisdiction until the end of the period, upon which the judge is replaced with a successor.

As noted above,\footnote{See TAN ___-___, supra.} one interesting consequence of viewing adjudication as a type of an empirical learning is that it lends itself to an endogenous definition of "precedent." According to this conceptualization, a prior holding has precedential weight when a later judge would choose rationally to follow it even when the facts of the case before her alone suggest the opposite outcome. Formally, in the context of this model I define a legal rule as precedent when it is common knowledge that subsequent judges will not deviate from that rule regardless of what they observe in the instant case.\footnote{This "predictive" element of precedent is consistent with the description often ascribed to legal realists. See Dworkin, supra note __, at 153.}
(b) The Emergence of a Precedential Cascade

Having set up a basic framework, it is now possible to describe how a legal precedent might emerge, as judges sequentially learn about the population they endeavor to regulate. As will become apparent, precedent (at least within this model) is likely to arise cataclysmically, as judges quickly gravitate to one of the two possible legal rules and tend to stay there. Nevertheless, there is no guarantee that the arrived-at precedent will correspond with each judge’s efficiency concerns.

Let us begin with the first-period judge, \( J_1 \), who hears her case having no good prior information about the characteristics of the victim population (i.e., \( \emptyset \)). With such limited information, the judge has very little to depend on other than the facts she observes in the first case, and thus such facts may play a pivotal role in determining her ultimate decision. Indeed, it is straightforward to demonstrate that the first judge will favor the pro-victim rule \( (y_1=1) \) whenever the first case involves a high-cost victim, and a pro-injurer rule \( (y_1=0) \) if the opposite is true. To be sure, the first judge is not entirely confident in issuing a holding based on such paltry knowledge; but from her perspective, limited knowledge is better than none. Note, however, that the initial judge’s holding is somewhat informative to her successors. Indeed, it signals precisely the type of victim she has observed.

Let us now move on to the second-period judge, \( J_2 \), who (for argument’s sake) inherits a history in which first judge has held for the pro-victim rule (indicating that \( J_1 \) observed a high-cost victim). Should \( J_2 \) also observe a case involving a high-cost victim, it is clear that he will similarly hold for the pro-victim rule \( (y_2=1) \). Indeed, in such a circumstance, \( J_2 \) enjoys the benefit of two proverbial draws from the urn, both of which have revealed a high-cost victim. Since the odds of such an event occurring are much higher when high-cost victims outnumber low-cost victims, the second judge’s observation will reinforce his confidence in the pro-victim rule. If, on the other hand, the second-period judge observes a case involving a low-cost victim, he will realize that, combined with his inferences about the first judge’s observation, he faces conflicting data about the population. To be sure, both of these data are

\[36\]See infra.
valuable to the second judge. However, they jointly indicate that the population of high- and low-risk victims is about equal, and he will be indifferent about whether to hold for the pro-victim or the pro-injurer rule. Because there may be good reasons for taking either route, let us assume that if $J_2$ faces such mixed signals, he will randomize between following her predecessor’s holding and issuing a holding suggested by the facts of the case before him.\(^{37}\) Nevertheless, note that the first-period judge’s holding does not have de facto compelling power over the second-period judge. In other words, because it is possible that $J_2$ will break from $J_1$’s holding, no precedent has yet emerged.

This conclusion may change with the third judge. In particular, suppose $J_3$ inherits a history in which both $J_1$ and $J_2$ have favored a pro-victim rule (i.e., $y_1 = y_2 = 1$). What might she infer from this history? Certainly, she can infer that the first judge must have observed a case involving a high-cost victim, or $J_1$ would have issued the opposite holding. Moreover, our third-period judge can infer that her immediate predecessor probably also observed a high-cost victim.\(^{38}\) This inference, as it turns out, is a strong one. It is so strong, in fact, that even if $J_3$ were to hear a case involving a low-cost victim, the judge—after accounting for the inferred revelations of her predecessors—would be disinclined to break from the rule announced by $J_1$ and ratified by $J_2$. This reasoning suggests that if the holdings of first two judges happen to be the same, a precedential cascade will emerge, and $J_3$ will rationally choose to follow that precedent, notwithstanding the facts of the case before her.

Although the discussion thus far has been limited to the first three judges, a precedential cascade may develop in later periods as well. For example, should the first two judges issue contradictory opinions, then the third-period judge will infer that her immediate predecessor must have observed a case whose facts differed from the first-period judge’s. Here, the split precedent she inherits does not send a very strong signal to $J_3$, and thus she will use the case before her to break the tie. In such a

\(^{37}\)This assumption is conventional in game theory, and can easily be relaxed.

\(^{38}\)Given what $J_1$ reveals about $\theta$, though it is possible that $J_2$ observed a low-cost victim case and randomized, this is less likely than the alternative that $J_2$ observed another high-cost victim.
situation, a cascade could still begin in a subsequent period. Indeed, it is possible to demonstrate that as soon as the historical record of holdings becomes sufficiently lop-sided in favor of either the pro-injurer or pro-victim rules, a precedential cascade will begin and will continue indefinitely.\(^{39}\)

(c) Significance

Although the process by which cascades occur in this framework is interesting in its own right, of greater import for current purposes is whether an emerging precedent will bear any resemblance to the normative goals of the judges who created it. Interestingly, the answer to this question turns out to be “not necessarily.” More precisely, when and if a precedential cascade begins, the resulting legal rule may—with high probability—vary from the efficient legal rule, even though each judge was assumed to be efficiency-minded and rational.

In order to illustrate this point more concretely, suppose that the true proportion of high-cost victims is equal to 40 percent, and thus the majority of victims (\(i.e.,\) 60 percent of them) are the lowest-cost risk avoiders. Clearly, the most efficient legal rule in this circumstance—at least on average—is the pro-injurer rule, which places all the risk on the victim (regardless of his individual avoidance cost). Nevertheless, the there is no guarantee that a stable precedent (when it emerges) will coincide with the most efficient rule. Indeed, it is possible to specify a number of situations in which a cascade is built up around the less efficient doctrine. For example, should the first two judges both hold in favor of the pro-victim rule (a contingency that occurs 28% of the time\(^{40}\)), a cascade will begin immediately in the next round. Even in later periods, an inefficient cascade is also possible.

---

\(^{39}\)See the appendix, \textit{infra}. Moreover, it can be shown that in this sort of framework, a precedential cascade will almost always occur. \textit{See id.}

\(^{40}\)Explicitly, such a history would will occur whenever the first judge sees a high-cost victim and either (1) the second judge sees the same; or (2) the second judge observes a low-cost victim, randomized between the two rules, and ends up selecting the pro-victim rule.
Figure 1 illustrates the probability (measured from the *ex ante* perspective) that the jurisdiction’s prevailing rule will be inefficient during each successive period. Note that during the initial periods, the probability of an inefficient rule rapidly falls. This seemingly convergent process quickly slows in subsequent periods, however, as the likelihood that a cascade has already begun increases. In fact, the probability that the jurisdiction adopts an inefficient never falls below 36.84 percent, *no matter how many periods one considers.*

![Graph showing probability of inefficient rule over periods](image)

**Figure 1**: *Ex Ante* Probability of an Inefficient Rule Each Period

A few features from this analysis bear immediate emphasis. First, note that stable precedents tend to emerge rather rapidly in this example. Indeed, in the case depicted in Figure 1, there is approximately a 76-percent chance a stable precedent will have emerged by the third period, growing to 98-percent by the seventh period. Nevertheless, whenever a cascade starts, *it is will be the wrong one nearly two-fifths of the time.* Consequently, a rapid convergence to a stable precedent reveals little about whether or not the resulting rule corresponds to an efficient outcome, making it difficult for an outside observer (such as a legislature, judge or legal scholar) to infer very much from the fact that a “watershed” holding is repeatedly affirmed. Equivalently, one can never be sure that a stable rule has emerged because “its factual
premises have been . . . validated by repeated testing,"\textsuperscript{41} or rather because of a chance dependence on initial cases that are statistical outliers.\textsuperscript{42}

Second, it is important to note that once a precedential cascade begins, there is nothing within the internal framework of the model to snuff it out. Indeed, $J_4$, knowing that $J_3$ rationally emulated her predecessors, is essentially the same position, and would therefore also find it optimal to follow suit. And so in turn must $J_5$, $J_6$, and all who follow. This point motivates related observation: when a precedential cascade begins, the process of judicial learning essentially stagnates. Indeed, consider the dilemma of $J_4$, who knows that her predecessor ($J_3$) was caught up in a cascade. The fact that $J_3$ chose to follow precedent signals nothing about whether her case involved a low- or high-cost plaintiff (indeed, $J_3$ would have issued the same holding regardless of what she observed). By logical implication, then, succeeding judges can learn nothing about the facts observed by their predecessors after $J_1$ and $J_2$. The marginal judge, no matter where she falls in the subsequent sequence of decision makers, has no more confidence than did $J_3$ that the holding she issues is the right one (even if it ultimately turns out to be).\textsuperscript{43}

Third, although a cascade (once commenced) is quite stable, it is also quite brittle. All sequential actors in a cascade realize that the group’s behavior is extremely path-dependent on the first few cases. Consequently, if an external shock from outside the above framework (caused by some public disclosure of information, or even population drift) were to occur, it might easily “break” the cascade.\textsuperscript{44}

Finally, while the example above illustrates how precedents may diverge from efficiency goals, the implications of the herding approach—at least in theory—need

\textsuperscript{41}See Posner, supra note ____.

\textsuperscript{42}Easterbrook (1982) lists the lack of path-dependence as a desideratum of a common-law system.

\textsuperscript{43}Some commentators seem to overlook this point. See, e.g., Kuran & Sunstein, supra note 9, at 722 (“In the course of an informational cascade, the perceived validity of a claim grows progressively stronger with the number of apparent believers, and people’s doubts weaken, possibly even disappear).

\textsuperscript{44}I take this issue up more explicitly infra at TAN __-__.
not be limited to a single motivational account of judging. Indeed, suppose, for example, that judges were uninterested in efficiency, but rather they were motivated by a desire to remedy distributional inequities within the population. Suppose further that in pursuit of this goal, judges wished to manipulate the contours of tort law (such as the calculation of damages) so as to effect transfers from the more well-off to the less well-off in society. Clearly, such equity-minded judges would be interested in gleaning information about whether, in the main, victims or injurers tended to be the most “well off” in a distributional sense. So long as there is some heterogeneity in the population of cases (i.e., some cases involve tortfeasors who are better off than their victims, and some involve the opposite), the argument developed above would seem to apply with equal force. Early opinions might reflect the content of what those judges observed—sending signals that effectively bind successors. Indeed, as noted in the introduction, a cascade theory of precedent—if plausible—holds important implications for virtually any positive account of law (efficiency-oriented or otherwise) that implicates some component of judicial learning.

II. ARE PRECEDENTIAL CASCADES PLAUSIBLE?

To be sure, the possibility of precedential cascades poses some unsettling questions for legal theory and doctrine. Just how alarming these questions are, however, is a different matter—one that ultimately turns on whether, all things considered, the conditions that give rise to herding behavior are a plausible description of the judicial process. It is to this matter I now attend, focusing on six necessary conditions for “bad” precedential cascades to occur: Rule-boundedness, decisional opacity, judicial homogeneity, short judicial tenures, flat hierarchies, and population

---

45Some law-and-economics scholars have argued that pursuit of distributional goals should be the province of the tax system alone, and that non-tax legal rules (such as tort doctrines) should not be used as an instrument for redistribution. See Louis Kaplow & Steven Shavell, Taxes vs. Legal Rules? An Economic Analysis, ___ J. Legal Stud. ___ (1994). This hypothesis has recently been challenged however. See Chris Sanchirico, Taxes vs. Legal Rules: A Reply to Kaplow and Shavell (USC working paper, 1998) (demonstrating that Kaplow & Shavell’s argument depends on an unrealistic critical assumption).

stationarity. I argue that not only is each of these conditions far from ubiquitous when viewed in isolation, but that their mutual confluence is especially rare. Moreover, a number of these conditions are easily avoidable by the prudent design of a judicial system, manifestations of which are already largely present within the current institutional regime.47

(a) Rule-boundedness.

For a precedential cascade to occur (or any other type of cascade, for that matter), it is necessary that judges be “rule-bound.” In other words, the set of legal outcomes that are available to judges must be either binary or must involve a small and finite number of outcomes. The example in Part I quite obviously involved rule-bound behavior: each judge was constrained to select from one of two possible rules: strict liability (with compensatory damages), or no liability. There was no middle ground. It is this constraint that created the chief signaling problem for judges caught up in the cascade: were a judge to observe litigants that did not “match” the inherited rule, she was not free to fine-tune her holding, moving incrementally away from the status quo ante. Instead, if she wanted to abandon existing precedent, she had to opt for a dramatically opposed doctrine. In a cascade, a single contrary observation is (by definition) unable to motivate a judge to take such drastic measures. As a result, the information transmitted by a rule-bound judge to her successors contained so much noise as to have no informational value.48

47Although I will animate my discussion in this section using the explicit model developed in Part I, supra, the necessary conditions described below are generic to cascade models, and not confined to any particulars of the example illustrated above.

48In technical terms, the intuition is as follows: If each actor’s action space were sufficiently robust to allow for such incremental “fine tuning” by each successive judge, it will often be possible to “invert” the strategy that turns observations into actions, inferring with certainty the signal the judge must have seen. See generally In Ho Lee, On the Convergence of Informational Cascades, 61 J. Econ. Theory 395 (1993).
Undoubtedly, rule-bound judging may be an accurate description of some legal doctrines. Many pragmatists, such as Cass Sunstein, in fact favor casuistry over rule boundedness. Liability, for example, is often a binary outcome by definition. But in many practical situations, judges are not strictly rule-bound. Indeed, judges frequently have tremendous freedom to issue tailored holdings (by announcing standards rather than rules, for example), which in turn may allow for more accurate inferences to be made by judicial successors. Moreover, even when a judge is bound to a binary choice on one issue, she may have a wider berth on others. For instance, courts that decide on liability matters must often also announce applicable burdens of proof, legal presumptions, and remedies. Each of these ancillary decisions may represent a valuable source of information to later judges, even if the instant judge is unwilling to dislodge the liability rule. When incorporated into the example from Part I, then, such arguments tend to undermine the likelihood of cascades, enabling each judge to embed a credible signal of the facts she observed into a finely-tuned legal opinion. Thus, if a theory of judicial herding is to have any viability, it must be limited to cases in which judicial actors are rule-bound.

50Many pragmatists, such as Cass Sunstein, in fact favor casuistry over rule boundedness. See SUNSTEIN, LEGAL REASONING AND POLITICAL CONFLICT (1995) 136-47. On the other hand, Sunstein admits the necessity of rules in many cases, including default (and other “privately adaptable”) rules within contracts and corporations. Id. at 154-58.

50In the example from Part I, for instance, there was a logical reason for adjudication to be more rule-bound in its orientation. Because an injurer is frequently not in a position to determine herself whether the potential victim is a low cost risk avoider (i.e., potential victims are “observationally equivalent” to one another), she cannot discern whether the victim is a high-cost or low-cost individual, and this may give rise to a type of coordination failure where both or neither parties engage in precautions. In many other applications, the assumption of rule-boundedness seems perfectly natural as well. For instance, one straight-forward extension of this model deals with the formulation of the appropriate “default rules” for contract law: i.e., rules for interpretation of a contract when the parties have not specified their rights and duties in a particular situation. The normative theories about default rules specify that because the set of contracting parties is observationally equivalent before writing their contracts, an efficiency-minded court must try to formulate an optimal “off-the-rack” default. Much of corporations law also centers on the creation and judicial interpretation of such “off-the-rack” default rules. See Easterbrook & Fischel (1991).

51For an explanation of the relatively greater nuances standards, see Louis Kaplow, Rules versus Standards: An Economic Analysis, 42 DUKE L.J. 557 (1992) (arguing that standards can be decided ex post, and rules cannot).
(b) Decisional Opacity.

Another necessary condition for cascades is that communication between actors be limited. In stock-market applications of herding behavior, for example, market participants are able to observe transactions only, and not the motivating rationales the buyers and sellers.\textsuperscript{52} Similarly, a precedential cascade implicitly assumes that succeeding judges are able to observe only what predecessor courts have decided—but not \textit{why} they decide it.\textsuperscript{53} On its face, such an assumption within a judicial system is \textit{at least} a stretch, and at most demonstrably false. Indeed, all appellate level courts (and even some trial level courts) not only announce outcomes, but they also provide a written opinion describing both the facts of each case and how they reached their conclusions. Most importantly, the ability to write an opinion enables a judge who follows precedent in the face of contrary facts to signal her reluctance at doing so.

Allowing for written opinions would change the conclusions of the example in Part I dramatically. Indeed, consider a third-period judge in the earlier example who inherits a history of two pro-victim holdings, but who observes a case in which the injurer (rather than the victim) is the high-cost risk avoider. Clearly, this single observation may not be enough to induce $J_3$ to break with precedent, but a written opinion allows her both to follow her predecessors’ lead, \textit{and} to signal her ambivalence about doing so.\textsuperscript{54} Moreover, subsequent judges, in making inferences from existing case law, would tend to pay greater attention to $J_3$’s reasoning than her holding (which they know to be uninformative). Consequently, the judicial learning process would continue rather than stagnate at the third period. Figure 2 demonstrates how the inclusion of such written opinions would affect the likelihood of an inefficient cascade over the same interval of time in Figure 1. As the Figure

\textsuperscript{52}See Lee, \textit{supra} note ___.

\textsuperscript{53}See, \textit{e.g.}, Banerjee, \textit{supra} note ___, at 798.

\textsuperscript{54}See, \textit{e.g.}, Lake River Corp. v. Coarborundum, 769 F.2d 1284 (1985) (noting the lack of theoretical and empirical support for the liquidated damages rule, but concluding: “however this may be, we must be on guard to avoid importing our own ideas of sound public policy into an area where our proper judicial role is more than usually deferential”);
illustrates, judicial learning will not stagnate as it did earlier when written opinions are available. Indeed, if one were to let this process continue indefinitely, the prevailing legal rule will converge (with probability one) to the most efficient legal rule.

It is, of course, debatable whether written opinions are always perfectly candid reflections of a judge’s thoughts. Indeed, the legal realists first argued that judicial opinions are rarely sincere, reflecting little more than stealthily-cloaked social theory in the garb of formal legal precepts. Moreover, recent “pragmatist” theorists have even endorsed the idea of opaque decisions. Cass Sunstein, for example, has argued (somewhat ironically) that judges should be somewhat insincere (or at least

\[55\text{See, e.g., }\text{Thurman Arnold, The Symbols of Government (1935); Max Radin, The Theory of Judicial Decision: Or How Judges Think, 11 A.B.A. J. 357, 360 (1925); Charles Clark, Real Covenants and Other Interests Which Run with the Land (2nd ed. 1947) (noting that the “touch and concern” requirement has devolved into little more than an instrument for randomness); Felix Cohen, Transcendental Nonsense and the Functional Approach, 35 Columbia L. Rev. 809, 812 (1935) (making a similar argument about personal jurisdiction doctrine). Karl Llewelen’s now-classic article on judicial interpretation of statutes underscores this point, as Llewelen notes some twenty-eight separate pairs of “canons of construction,” each pair seeming to suggest opposite behavioral norms, such as the canon “Every word and clause must be given effect,” paired with its opposing canon, “If inadvertently inserted or if repugnant to the rest of the statute, [a word] may be rejected as surplusage.” K. Llewelen, _____ 3 VANDERBILT L. REV. 401-2 (1950).}

\[56\text{That Sunstein has advanced this argument is somewhat surprising, given that he is one of the principal contributors to the legal cascades literature, and a champion of institutional reforms that reduce the likelihood that “bad” cascades occur. See Kuran & Sunstein, supra note 9, at 762}
imprecise), confining the justificatory components of their holdings to “low-level” principles that are consistent with numerous normative theories of law. The rationale for such a practice, apparently, is to arbitrate and construct a type of consensus or “incompletely theorized agreement” between such competing meta-theories (or at least a number of them). Doing so, the argument asserts, is the principal means by which adjudication becomes viable as a type of social ordering in a pluralistic world. At the same time, however, such opacity comes at a distinct cost, for it tends to increase the likelihood that a bad cascade will occur.

Be that as it may, it seems unlikely that all (or even most) judicial opinions are hopelessly opaque. Sole presiding judges need not build a consensus among their counterparts (at least to reach an opinion), and they can therefore afford to be relatively more candid. Moreover, even among empaneled judges, the prospect of dissenting opinions, concurring opinions, opinions that concur only in the judgment, and the like all represent forms of information dissemination capable of obviating a future information cascade, even when a dispositive consensus has been struck. Consequently, a viable theory of precedential cascades must also be limited to situations where written opinions are sufficiently opaque.

(“A major challenge for any democratic system is to institute safeguards against harmful cascades”).

A related, though conceptually distinct pragmatist account of opacity asserts that judges may adhere to precedent without introspective analysis of the effects of the instant case as a means of avoiding distress and guilt over issuing difficult holdings. See Scott Altman, Beyond Candor, 89 Mich. L. Rev. 296, 305 (1990).

CASS SUNSTEIN, LEGAL REASONING AND POLITICAL CONFLICT (1995). To be fair, most of Sunstein’s argument is built around advocating opaque articulations of normative goals rather than facts. But nonetheless, there may be facts in a case that, if used in an opinion, betray the normative theory of the opinion writer. In such situations, an incompletely-theorized agreement might necessitate the omission of such facts, and the creation of a potential cascade. Consider, for instance, a judge who is attempting to build a consensus among other judges whose respective normative allegiances are split between efficiency and distributional concerns. Our judge may have access to a number of facts, including who is the lowest cost risk avoider (a factum that is critical to efficiency-minded concerns), and who between the injurer and the victim is the wealthiest (a factum that is critical to distributional concerns). The Sunstinean account (as I read it) suggests that this judge may prefer to issue a stark holding (e.g., “the injurer is liable”) rather than one that reveals an allegiance to one particular meta-theory or another (e.g., “the injurer is liable because she appeared to be the least-cost risk avoider”). By omitting such facts, this judge may be able to garner support from judges who pursue distinct (but in this case consequentially equivalent) norms.
(c) Judicial Homogeneity

A third necessary condition for a cascade is that the motivational objectives of all actors must be mutually similar and commonly known. In the model from Part I, for instance, it was common knowledge that all judges were identical, one-period efficiency maximizers. This knowledge enabled later judges to make inferences about early predecessors' cases based solely on the holdings in those cases. If judges do not share a similar motivational “fabric,” however, then it is much more difficult to make such inferences. Referring back to the model in Part I, suppose that the first two judges, \( J_1 \) and \( J_2 \), were interested solely in efficiency, while the third-period judge, \( J_3 \), was interested solely in distributional equity. Would a sequence of pro-victim holdings from \( J_1 \) and \( J_2 \) be sufficient to convince \( J_3 \) that she should simply follow suit? It seems unlikely. Unless the identity of the low-cost risk avoider were sufficiently correlated to the identity of the worst-off litigant, then the first two holdings are of relatively little value to \( J_3 \) when she makes her decision. Instead, she would place significantly greater decisional weight on the facts revealed by case before her, thereby breaking the possible cascade.\(^{59}\) This dampening effect is even stronger when judges are unsure of each other’s motivations. Asymmetric information of this sort magnifies

\(^{59}\)Kuran & Sunstein, supra note 9, appear to posit the opposite of this claim. See id at 722 (noting that a “precondition of any cascade [is] heterogeneity in individual responsiveness to social signals”). This claim is inconsistent with the arguments made in much of the cascade literature. See sources in note _, supra. It appears that their claim has more to do with the question of whether a cascade will start at once or build slowly, as less responsive individuals slowly join the bandwagon. It is important to note, however, that if bandwagon effects occur among heterogeneous individuals in the manner described by Kuran and Sunstein, it is unlikely to be a problematic cascade. Indeed, less responsive individuals are (by definition) less likely to infer wisdom from their predecessors in action, and will pay more attention to their own private signals. In this sense, the fact that a relatively unresponsive individual joined the herd signals valuable new information to her successors.

Heterogeneity, however, may play an interesting role in reputational theories of judging. Some herding theories, for instance, posit that individuals of different “skill levels” may prefer to follow the crowd if they are concerned about being exposed as of low-quality. Jeff Zwiebel in fact argues that in such situations, either the very talented or the very untalented actors may be willing to break with the herd. Zwiebel, supra note 4. Note, however, that even within this framework, heterogeneity makes monolithic bandwagon effects less rather than more likely.
the incentive to place one’s own observation above the (imperfectly revealed) observations of others.\textsuperscript{60}

\textit{(d) Short Judicial Tenures.}

Herding behavior, in a rather interesting twist of irony, bears a striking resemblance to the so-called “tragedy of the commons.”\textsuperscript{61} Recall that the tragedy of the commons occurs when self-interested shepherds, disregarding the welfare of others, have inefficient incentives to over-graze the commons, thereby imposing a negative externality on one another. A similar form of externality infects an information cascade, though one that exists over time rather than space. Cascades occur because individual decision-makers seek to maximize their own benefit by following the herd, disregarding the welfare of future decision-makers, who would benefit from learning about predecessors’ observations.

One proposed solution to the tragedy of the commons is to vest ownership of the commons in a single shepherd, who could control the rate of grazing.\textsuperscript{62} Since the new owner must bear the full costs of over-grazing, she will tend to internalize the externality, and—at least in theory—the inefficiency will disappear. The analogous solution within information cascades would be to allow a single actor to make decisions over numerous periods. In the extreme case, if a repeat actor were infinitely-lived, she could never be caught in a cascade, since she would have personal recollection of the signals she observed in each period before.

\textsuperscript{60}Moreover, even in those situations where judges share common goals, a cascade need not always emerge. If, for example, all judges were committed to a goal of \textit{dynamic} rather than \textit{static} efficiency, then in each period a judge might be willing to ignore existing precedent and base his holding solely on the facts before him. Although this approach “throws away” information that could be useful today, it allows the instant judge to provide more data for tomorrow’s judges.

\textsuperscript{61}See Hardin (cite).

\textsuperscript{62}\textit{Id.} at __.
As it turns out, common-law judicial systems have an excellent device for ensuring repeat play: long judicial tenures. All Article III judges have life tenure, as do a number of state judges. The remainder frequently have generous terms of office. Allowing judges to have long tenures enables them to hear a number of similarly-situated cases, updating their beliefs over time with each succeeding case. Consequently, long tenures (even if not infinite) can mitigate the problem of herding behavior significantly. If, in the example from Part I, a single judge served for three periods rather than one, then the probability of an efficient precedent eventually emerging would increase from (approximately) 63 percent to 78 percent, and the corresponding probability that an inefficient precedent emerges would decrease from 37 percent to 22 percent.

Thus, a theory of precedential cascade is far less compelling when individual judges serve out long terms on the bench.

---


64 Although not discussed in the text, one can replicate the beneficial effect of long tenures by allowing judges access to information that goes beyond current or historical cases before the court, such as statistical studies of industry practice or population base rates. Interestingly, however, numerous procedural constraints in American courts constrain judges to confine their opinions principally to precedent and the facts of the case before them. The most notable of these are Constitutional constraints, requiring for instance that courts may not exercise jurisdiction unless necessary to resolve “cases” and “controversies.” This non-investigatory capacity of the court is suggestive that courts are not active researchers by nature, but rather are passive adjudicators. They do not conduct systematic studies to determine the effectiveness of the legal rules they promulgate, and in fact often disavow such a role, claiming it (perhaps correctly) to be one that is more fit for a legislative body. See, e.g., U.S. v. Topco, 405 U.S. 596, ___ (stating, “To analyze, interpret, and evaluate the myriad of competing concerns and the endless data . . . the judgment of the representatives of the[Congress] is required.”). Additionally, the constitutional constraints by such doctrines as the mootness, ripeness, and the political and administrative question doctrines prevent courts from inquiring into cases that cease to be in dispute, are not yet in dispute, or are in some way beyond the competence of a court as a passive actor. See 13 Charles Wright et al, Federal Practice and Procedure ___ (1984). Each of these doctrines constrains the ability of a court to act in a capacity as a social policy researcher. Other procedural doctrines, such as the relevancy rules of evidence may further impede a court’s access to large “samples” from which to learn. Fed. R. Evid. 401.
(e) Flat Hierarchies.

Another important assumption from the previous section was that the hierarchy of courts was relatively flat. In particular, there was only one court in each period, whose holding was not subject to review by a higher court. If one were to add a hierarchical judicial system to the example from Part I, the likelihood of a “bad” precedential cascade might be significantly mitigated. Indeed, on an extremely general level, the appeals process can be seen as a form of error correction device, particularly in situations where a cascade has begun or seems likely to begin. Appellate judges, by virtue of their more prestigious positions, are generally perceived to be more skilled at adjudication than are their counterparts lower down in the hierarchy. By reviewing a lower-court opinion (or a set of them at once), the appellate judge may be able to extract a more informative signal than did her lower-court counterparts, breaking a destructive information cascade if such an action is warranted. Viewed in this sense, then, the hierarchical structure of the judicial system places an important check on the viability of run-away herding effects. Although a cascade may still be possible even in a hierarchical system, it is substantially less likely than in a non-hierarchical one.

(f) Population Stationarity.

Finally, for an information cascade to have any durability, there must not be significant “drift” in the underlying empirical characteristics of interest. In the model from Part I, for example, the absence of drift implies that the population of injurers and victims—and particularly the ratio of high-cost to low-cost victims—must remain relatively constant over time. For if it did not, then it would be difficult for late-

---


66Indeed, in their analysis of cascades, Daughety and Reinganum analyze a set of circuit court cases in which a cascade may have occurred, but one that was nonetheless broken by the Supreme Court. Daughety and Reinganum, supra note __, at 4 et seq.

67Interestingly, the Supreme Court has a practice during deliberations of taking straw polls in reverse order of seniority. If one believes that tenure on the court is a good proxy for signal quality, then such an institution would constitute a sensible anti-herding device.
moving judges to infer much at all about early judicial opinions. Indeed, such opinions might have been statistically justifiable at the time, but need not if the relevant population has undergone considerable change.

It is important to note that the complete absence of drift is not necessary for a cascade to form. Indeed, so long as the population changes at a sufficiently slow rate, brief cascades could emerge. However, such cascades would be relatively infrequent and brittle, breaking easily once the drift of the population had out-paced the wisdom of early opinions. As such, a lack of stationarity makes the prospect of cascades less troublesome.

Each of the above six conditions is necessary for a theory of precedential cascades to be both viable and significant. None of them, however, when viewed alone, is likely to be ubiquitous in practice. More important, however, is the necessity that these conditions must occur jointly, and not alternatively. Hence, if there is a sub-universe of cases in which courts manifest a dangerous form of herding behavior, it is most likely limited to those situations in which courts are rule-bound, and judicial opinions are opaque, and judges are similarly motivated, and judges serve short (or ephemeral) terms, and appeal is unlikely, and population drift is relatively small. Perhaps such a sub-universe of cases exists; but if it does, it seems unlikely to represent a significant fraction of cases in most areas of litigation.

III. ARE PRECEDENTIAL CASCADES VERIFIABLE?

Although I have argued that herding behavior is implausible as a general account of judging, its contributions could still have considerable value if appropriately limited to those situations where the dangers of such phenomena appear large. Indeed, if it were possible to isolate a strand of legal doctrine caught within a judicial cascade, it might be possible to employ reform efforts either to break the cascade (if it is a bad one) or validate it (if it is a good one).68 In substantive areas of

68See, e.g., Kuran & Sunstein, supra note 9, at 688.
law where cascades are problematic, for example, we might place greater requirements on written opinions, lengthen the tenure of judicial actors, build steeper judicial hierarchies, or encourage other forms of public information disclosure.

Each of these remedial efforts, however, turns on whether one can detect the emergence of a precedential cascade in specific legal contexts. Here it seems that two types of evidence might be of some assistance. First (and preferably), one might rely on aggregate court data, measured either cross-sectionally or over time, to help detect when a cascade has emerged. In the absence of such statistical proof, however, it might be possible to use anecdotal evidence to demonstrate the existence of a herding problem. Unfortunately, both of these approaches are frequently riddled with difficulties so profound that it is often impossible to determine whether an episode of apparent group conformity stems from a cascade or some other, more benign explanation.

(a) Empirical Evidence

Consider first the task of attempting to infer a judicial herding episode statistically, from aggregate cross-sectional or time-series data. Casual observations from such data are in fact what motivated much of the initial herding literature, including accounts of asset price bubbles, corporate conservatism, bank runs, political revolutions, medical therapies, and zoology. Each of these applications is typified by a trend that begins locally, is subject to rapid, serial emulation by other agents who disregard their own private information, and seems to hinge on one or

---


70E.g., Zwiebel, supra note 4.

71E.g., Chen, Yehning (1995).


74Galef (1976).
two initial movers who appear to trigger the cascade and its subsequent aggregate effects. Thus, one might ask, can similar phenomena be detected in legal doctrine? If so, can the quick spread of the doctrine be fairly interpreted as a precedential cascade?

Twentieth-century American jurisprudence is not lacking in rapidly-spreading doctrines. A particularly good example of such a trend the spread of the contract doctrine of unconscionability during last forty years. The doctrine, in its modern form at least, allows a party to void a contractual term or terms if it appears to the court that the term(s) is manifestly unfair (often called “substantive” unconscionability) and/or the term was the result of a bargaining process that gave such a party no meaningful choice (often called “procedural” unconscionability). Although a number of opinions dating as far back as the nineteenth century had invoked the term “unconscionable” while refusing to enforce a contract on public policy grounds, New Jersey was likely the first to adopt unconscionability as a separate, generally-applicable doctrine in 1960.\(^75\) The doctrine soon began to spread to other jurisdictions, many of which cited the early movers in almost a cursory fashion.

\(^75\) Henningsen v. Bloomfiled Motors, 161 A.2d 69 (1960). One of the first articulations of the doctrine is in the Uniform Commercial Code § 2-302, drafted during the 1950s. Though the UCC constitutes compelling authority only for cases of goods, the adoption of the doctrine described here was much broader than that, and applied UCC § 2-302 by analogy to other cases as well.
The swiftness of the unconscionability trend was relatively dramatic, though not simultaneous, as depicted in Figure 3. In 1970, the doctrine had been adopted by at most six states. By 1985, however, fully forty-one states had embraced the doctrine. Since then, the trend has obviously slowed, and only five more states adopted the doctrine.\textsuperscript{76} Like many other purported examples of herding, the movement toward unconscionability began in the Northeastern and New England states, spreading geographically and moving into the Southern states last.\textsuperscript{77}
Does this evidence suggest that the doctrine of unconscionability is a form of information cascade? It may. Indeed, a number of early decisions appear to cite other states’ cases liberally, indicating at least some informational interdependence between jurisdictions. At the same time, however, one of the largest difficulties one faces with rational herding models is that such phenomena are difficult to verify empirically in the face of other plausible hypotheses. Indeed, it may be that the detected trend is an artifact of some population-wide informational “shock” eliciting rapid cross-sectional adjustment, but one that outsiders might fail to measure or detect. For example, the timing of many of the “watershed” unconscionability cases (such as Williams v. Walker-Thomas Furniture) coincided with the publication of numerous high-profile empirical studies of poverty in the United States—tending to show that the relative plight of American poor was worse than many had thought. The timing also coincided with the arrival of a number of Left-Democrat-appointed

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980 Per Capita Income</td>
<td>-.00024</td>
<td>.00037</td>
<td>-.64817</td>
</tr>
<tr>
<td>1980 Poverty Rate</td>
<td>.11434</td>
<td>.12830</td>
<td>.89121</td>
</tr>
<tr>
<td>Southern State (dummy)</td>
<td>-1.48058</td>
<td>.91535</td>
<td>-1.61751*</td>
</tr>
<tr>
<td>1980 College Degree Rate</td>
<td>.25315</td>
<td>.19211</td>
<td>1.31770</td>
</tr>
<tr>
<td>Constant</td>
<td>-.65504</td>
<td>4.26831</td>
<td>-.15347</td>
</tr>
</tbody>
</table>

* = p < 0.10; ** = p < 0.05

N = 51 (50 states & District of Columbia)


The only independent variable showing weak significance is membership in southern states being weakly predictive of failure to adopt. Further it is hard to know what to make of this variable, since it could be a geographic artifact of herding effects itself. The qualitative results reported above persist with a number of other permutations of the model (in which one or more of the independent variables are excluded). Moreover, the demographic variables do not have a statistically significant joint effect on pre-1986 adoption. $\chi^2 (46) = 47.346; p=0.417$.

78Cite.

79350 F.2d 445 (1965).

80Cites.
and elected judges, many of whom would likely be favorably disposed to
distributionally-minded doctrines, and unconvinced by the holdings of more
conservative judges that preceded them.\textsuperscript{81} These alternative phenomena are neither
byproducts nor causes of cascades, but either could give rise the identical trend noted
above.

This statistical identification problem is significant, particularly for those
interested in normative legal reform. Indeed, it is both risky and imprudent to
advocate significant reform measures on the basis of empirical observations that lend
support to any number of plausible interpretations, only some of which are cause for
alarm. A sincere belief in the prevalence of cascades is insufficient. Moreover, while
the discussion above is limited to a particular doctrinal candidate for precedential
cascades (\textit{i.e.}, the unconscionability doctrine), similar problems of empirical
verifiability infect other endeavors as well.\textsuperscript{82} At core, the empirical challenges are
quite general: because herding phenomena are the handiwork information asymmetries
that cannot be directly observed by outsiders, it becomes extremely difficult to
demonstrate that a herding hypothesis is empirically more likely than any other
plausible alternative, which is itself either difficult to observe or has been erroneously
excluded from the analysis. Consequently, most attempts to verify the existence of
herding effects in empirical data are subject to extreme hazards.\textsuperscript{83}

\textit{(b) Anecdotal Evidence}

Although empirical data seems unlikely to provide uncontroverted proof of the
prevalence of legal cascades, it may be possible to use anecdotal evidence to suggest
that cascade effects may have occurred in some isolated situations. Unlike the
aggregated statistical approach above, an anecdotal approach places greater emphasis

\textsuperscript{81}Indeed, Skelly Wright, the author of \textit{Williams v. Walker Thomas}, was himself a Democrat
appointee (Truman) to the federal bench, having been elevated to the DC circuit by President
Kennedy in 1962.

\textsuperscript{82}Cites.

\textsuperscript{83}Matters may be different with experimental data, however. See Welch et al (1999).
on determining whether a series of judicial opinions constituting a possible cascade reflect (1) an understanding of precedent as a reflection of judicial learning; and (2) an unsubstantiated confidence in that inherited learning that tends to quell critical analysis of the instant facts in a case before the judge.

A few advocates of cascade theory have offered anecdotal examples of possible episodes of judicial herding. Andrew Daughety and Jennifer Reinganum, for instance, posit that a series of six similarly-situated opinions in the First, Second, Third, Fourth, Sixth and Seventh circuits preceding the recent U.S. Supreme Court opinion in Eastern Enterprises v. Apfel\(^{84}\) might have constituted a precedential cascade.\(^{85}\) Each of the appellate-level cases upholds the constitutionality of a basic provision in the Coal Industry Retiree Health Benefit Act of 1992, in the face of both takings and due process challenges.\(^{86}\) Moreover, the series of cases appears to satisfy both of the above criteria. Each cites to all its predecessors for persuasive authority, and the opinions appear to become more abbreviated. Indeed, by the time Eastern Enterprises is litigated in 1997, the First Circuit Court of Appeals noted, “The constitutional arguments [by the appellants challenging the Coal Act] are retreads which have taken their lumps from circuit courts of appeals in five other circuits,” and “[a]lthough these decisions are not binding on us, we find them convincing.”\(^{87}\)

Do the cases that precede Eastern Enterprises constitute a horizontal cascade? If so, is it something we should be concerned about? Both of these are difficult questions to answer. As to the first, it is true that each of the cases cited to its

---


\(^{85}\)Daughety & Reinganum, supra note __, at 5-7.


\(^{87}\)Cite.
predecessors for support. However, none of them is lacking in long and even protracted analysis. (The 1st Circuit’s opinion in Eastern Enterprises, for example, is nearly thirteen two-column pages long). Moreover, one can find examples in each case of factual analysis that belies a herding hypothesis. In a number of cases, the written opinions delve much further into the existing case law than a simple catalog of previous holdings, often distinguishing contrary authority on the basis of distinct facts. As noted above, a pre-condition for a cascade is that such facts cannot be communicated among judges outside of their rule-bound holdings. Finally, even if one were convinced that the cases described represent a type of judicial cascade, there may be little cause for alarm. Indeed, within three years of the initial appellate court opinion on the issue, the Supreme Court (arguably in possession of a more accurate signal) struck down the regulation as unconstitutional. As noted above, the possibility of an appeals process tends to mitigate the severity of a precedential cascade—and in this case it appears to have worked as designed.

It goes without saying that legal discourse often revolves around the distinct effects of “landmark” cases, which can, in turn, create a landslide of similar holdings. Brown. Plessey. Roe. Lochner. All of these cases have had a perceived effect of “locking in” subsequent courts on the issue involved. Perhaps as a result, considerable hand-wringing tends to permeate these areas of law, over whether such cases represent the correct normative trajectory. However, those interested in legal reform should be wary of dressing these doctrines up in the ill-fitting garb of cascades. Doing so may give rise to even less defensible policy outcomes.

---

88 E.g., Eastern Enterprises, 110 F.3d at 161 (“Unity Real Estate, on its facts, see 889 F. Supp. at 829-31, is easily distinguishable from the case at bar”).

CONCLUSION

Herding behavior has become an increasingly popular trope among legal academics in recent years. A recent LEXIS search finds 66 published law-review articles that refer to either “herding behavior” or “information cascades” or both since 1992. This trend is hardly surprising: traditional rational actor models have—in many cases—done a poor job of predicting individual behavior in practice, and are even worse at predicting group behavior. Asymmetric information models, such as cascade theory, offer a more attractive explanation of how individuals interact—one that may (at least in some contexts) be harmonious with behavioral pathologies that psychologists have observed for some time. Recent efforts to merge the two ideas into an explanation of legal doctrine present some thought-provoking ideas, but suffer from potentially severe problems with both plausibility and verifiability. They do not—at present—provide general and reliable account of legal evolution.

This does not mean, however, that considerations of herding behavior are wholly irrelevant to advance our theory of judging and judges. Indeed, cognitive biases or reputational concerns alone might drive judges to fall into a cascade. But if that be so, the case must be made on those terms, and not as a set of phenomena that exacerbate existing informational cascades. Alternatively, lawyers, litigants, or potential litigants may be significantly more susceptible to “bad” herding phenomena. Both judges and legal scholars should be aware of such possibilities, particularly in attempting to understand (and perhaps to reform) doctrines that hinge

---

90 A recent LEXIS search finds 66 published law-review articles that refer to either “herding behavior” or “information cascades” or both since 1992.

91 E.g., Kuran & Sunstein, supra note 9. Moreover, if one wishes to make a case for cascades solely on cognitive grounds, it is still necessary to test the hypothesis of a herding effect against other plausible hypotheses that would lead to unified group behavior (see above), and to account for other plausible cognitive pathologies that tend to diminish (rather than amplify) the likelihood of a cascade. See, e.g., Antonio Bernardo & Ivo Welch, On the Evolution of Overconfidence and Entrepreneurs, UCLA Anderson Graduate School of Management Working Paper #9-97 (1998) (arguing that overconfidence among a segment of a population can have beneficial herd-breaking effects).

on existing social norms or other aggregated behavior. But in pursuing such ends, it is important to be sure that it is not we—rather than those we study—who are the ones caught up in a cascade.

93Contract interpretation doctrine, for instance, often hinges on established trade custom. See, e.g., UCC § 2-202.
This Appendix contains a more formal analysis of some of the central intuitions developed in the body of the paper. I employ the following notation:

\[ J_i = \text{Judge in period } i \]
\[ c = \text{Injurer’s cost of avoidance (constant for all injurers)} \]
\[ x_i = \text{Signal received by } J_i \text{ about victim (1 if high-cost, 0 else).} \]
\[ c + d(x_i) = \text{Randomly-drawn victim’s cost of avoidance (} d(1) = D > 0 \text{ if victim is high-cost, and } d(0) = -D < 0 \text{ if victim is low-cost).} \]
\[ y_i = \text{ } J_i \text{’s announced holding (1 if pro-victim, 0 if pro-injurer).} \]
\[ \theta = \text{True proportion of high-cost victims (constant over time)} \]

The following Lemma describes how an efficiency-minded judge will behave with a given inherited history and given observation. This lemma applies regardless of whether each judge has access to no history, only a set of holdings, or a set of holdings and reasonings.

**Lemma 1:** The marginal judge \( i \) facing history \( h_i \) and observing facts \( x_i \) strictly prefers holding \( y_i = 1 \) over \( y_i = 0 \) if \( f(\theta | X_i, h_i) \) dominates the \( f(1-\theta | X_i, h_i) \) in the sense of first order stochastic dominance.

**Proof:** Given history \( h_i \) and observing facts \( x_i \), \( J_i \) strictly prefers holding \( y_i = 1 \) to \( y_i = 0 \) whenever the expected precaution costs are lower for \( y_i = 1 \) than for \( y_i = 0 \). Equivalently, \( J_i \) prefers holding \( y_i = 1 \) when:

\[ c < \int \left[ (c + D) \cdot \theta + (c - D) \cdot (1 - \theta) \right] dF(\theta | h_i, x_i) \]

Which, after simplification, is equivalent to:

\[ \int_0^1 (\theta) dF(\theta | h_i, X_i) < \int_0^1 (1 - \theta) dF((1 - \theta) | h_i, X_i) \]
The above expression will always be satisfied under first-order stochastic dominance; \textit{i.e.}, \( F(\theta \mid X_i, h_i) < F(1-\theta \mid X_i, h_i) \) for all \( \theta \in [0,1] \).

Using Lemma 1, it is possible to derive the following propositions:

\textbf{Proposition 1:} Suppose a judge had access neither to previous holdings nor to the private signals of previous judges (i.e., \( h_i = \{ \} \)), \( J_i \) will hold \( y_i = 1 \) if and only if \( x_i = 1 \).

\textbf{Proof:} “If”: Under myopia, where \( h_i = \{ \} \), every judge \( J_i \) is in the identical position as \( J_1 \). Thus, one need only demonstrate that \( J_1 \) will hold \( y=1 \) if and only if she receives a signal of \( x_1 = 1 \). Suppose \( J_1 \) does receive such a signal. The resulting posterior density she places on \( \theta \) (recalling the uniform prior) will then be:

\[
\begin{align*}
\frac{f(\theta \mid X_1 = 1)}{f(1-\theta \mid X_1 = 1)} &= \frac{\int \Pr(X_1 = 1 \mid \theta) \cdot f(\theta) \, \theta \, d\theta}{\int \Pr(X_1 = 1 \mid \theta) \cdot f(\theta) \, d\theta} = \frac{\theta}{2} \\
&= \frac{2\theta}{2}
\end{align*}
\]

Note that \( f(1-\theta \mid x_1 = 1) = 2(1-\theta) \), and thus \( f(\theta \mid x_1 = 1) \) must cross \( f(1-\theta \mid x_1 = 1) \) once from below. Consequently, \( F(1-\theta \mid x_1 = 1) > F(\theta \mid x_1 = 1) \). “Only if”: assume the judge receives a signal of \( X = 0 \). It is easily shown that the posterior distribution on \( f(\theta \mid x_1 = 0) = 2(1-\theta) \). The rest of the proof proceeds in an identical fashion as that above.

\textbf{Proposition 2:} Suppose instead that judge \( J_i \) has access to both previous holdings and the factual record that drove those opinions (i.e., \( h_i = \{ y_1, y_2, \ldots, y_{i-1}; x_1, x_2, \ldots, x_{i-1} \} \)). Judge \( J_i \) will prefer \( y_i = 1 \) (\( y_i = 0 \)) whenever the total number of \( x=1 \) signals exceeds (is less than) the number of \( x=0 \) signals.

\textbf{Proof:} Note first that if \( J_i \) has access to the factual records, she has all relevant information and may disregard the actual holdings of her predecessors. Define \( n_0 \) as the number of \( x=0 \) signals and \( n_1 \) as the number of \( x=1 \) signals received up to and including \( x_i \). The posterior distribution of \( \theta \) given \( n_0 \) and \( n_1 \) is given by Bayes rule as follows:
The case of $y_{i-1} = y_i = 0$ is identical, and is therefore ignored.

From the above expressions it is clear that $f(\theta \mid n_0, n_1) > f(1-\theta \mid n_0, n_1)$ whenever

$$\int_{\theta}^{n_0} \Pr (n_0, n_1 \mid \theta) d\theta > \int_{\theta}^{n_0} \Pr (n_0, n_1 \mid 1-\theta) d\theta$$

This implies that when $n_1 > n_0$, $f(\theta \mid n_0, n_1)$ crosses $f(1-\theta \mid n_0, n_1)$ from below but one time, and satisfies the single crossing property of Milgrom (1981). This ordering is reversed whenever $n_1 > n_0$. As such, $F(\theta \mid n_0, n_1)$ dominates $F(1-\theta \mid n_0, n_1)$ in the first order stochastic sense whenever $n_1 > n_0$, and vice versa.

Finally, consider the case in which judges only have access to prior holdings, but not prior opinions.

**Proposition 3:** Suppose historical holdings are opaque (i.e., $h_i = \{y_1, y_2, y_3, \ldots, y_{i-1}\}$). Define $n_0$ as the number of $y=0$ holdings and $n_1$ as the number of $y=1$ holdings through stage $i$. At the first stage $i$ in which $n_1 \geq n_0 + 2$ ($n_0 \geq n_1 + 2$), then for all $j \geq i$, $y_j = 1$ ($y_j = 0$), regardless of $x_j$.

**Proof:** Suppose that $i$ is even, and that for every stage $k \leq (i-2)$, $|n_0 - n_1| < 2$. This implies that at stage $i-2$, $n_0 = n_1$. Because (by assumption) a cascade has yet to occur, the sequence $\{y_1, y_2, y_3, \ldots, y_{i-1}\}$ perfectly reveals the content of the signals $\{x_1, x_2, x_3, \ldots, x_{i-2}\}$: half $(i-2)/2$ courts received signal $x=0$ and half $(i-2)/2$ received signal $x=1$.

Without loss of generality, suppose that $y_{i-1} = 1$. By the same argument as above, $y_{i-1} = 1$ reveals that $x_{i-1} = 1$. However, the fact that $y_i = 1$ could reveal that $x_i = 1$, or it could reveal that $x_i = 0$ and that $J_i$ randomized between $y_i = 1$ and $y_i = 0$ with respective probabilities $[\alpha, 1-\alpha]$ (with $y_i = 1$ being the realization of that randomization process). As such, $y_i = 1$ would occur with probability $\theta + (1-\theta)\alpha$.

Now, suppose that in addition to this historical vector of holdings, $J_{i+1}$ receives the signal, $x_{i+1}$. To demonstrate the above proposition, it is sufficient to show that $y_{i+1}$

---

94The case of $y_{i+1} = y_i = 0$ is identical, and is therefore ignored.
= 1 even when \( x_{i+1} = 0 \).\(^{95}\) Thus, suppose \( x_{i+1} = 0 \). \( J_{i+1} \)'s posterior density on \( \theta \) is given by:

\[
    f(\theta \mid h_i, x_{i+1} = 0) = \frac{\left[ (1-\theta) \alpha \right] \cdot (1-\theta)^{i/2} \theta^{i/2}}{1 \int_{0}^{1} (\theta + (1-\theta) \alpha) \cdot (1-\theta)^{i/2} \theta^{i/2} d\theta} = K_i \cdot \left[ (1-\theta) \alpha \right] \cdot (1-\theta)^{i/2} \theta^{i/2},
\]

where \( K_i \) is a constant.\(^{96}\) Using the expression above, it is easily verified that \( f(\theta \mid h_p, x_i=0) < f(1-\theta \mid h_p, x_i=0) \) if and only if \( \theta < 1/2 \), thereby satisfying the single crossing property.\(^{97}\) As such, \( \theta \) dominates \((1-\theta)\) in the sense of first order stochastic dominance. Importantly, once this condition is verified for \( J_p \), \( J_{i+1} \) cannot infer anything from \( y_i \), and is therefore in the same position as \( y_i \). Thus, for any even \( i \), the proposition has been shown.

If \( i \) is odd, then it cannot be the case that \( n_i \geq n_{i+2} \) unless this was also true for \( J_{i+1} \). The argument above establishes that \( J_{i+1} \) must have been in a cascade, and by stationarity it is established for \( J_i \) as well.

\(^{95}\)Certainly, if one can show this result, then it is straightforward to show the weaker result that \( y_{i+1} = 1 \) if \( x_{i+1} = 1 \).

\(^{96}\)Explicitly, \( K_i \) is given by:

\[
    K_i = \frac{2^{i+2} \cdot \Gamma\left(\frac{i}{2} + \frac{3}{2}\right)}{(1+\alpha) \sqrt{\pi} \cdot \Gamma\left(\frac{i}{2} + 1\right)}
\]

\(^{97}\)See Milgrom (1981).