
PRECEDENTIAL CASCADES: AN APPRAISAL

ERIC TALLEY*

INTRODUCTION

About a half century ago, a handful of social scientists began to formalize what was to become the analytical heart of neoclassical economics.¹ Under the broad rubric of “general equilibrium theory,” these scholars demonstrated (in varying degrees of mathematical sophistication) the longstanding intuition behind the so-called “invisible hand”: that is, that competitive markets could convert apparent disarray and fragmentation into order and harmony.² 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.³ An equally powerful corollary attended this central insight: that the prices that emerge in such markets

* Associate Professor, University of Southern California Law School. Thanks to Scott Altman, Jennifer Arlen, Ian Ayres, Antonio Bernardo, Erwin Chemerinsky, Marshall Cohen, David Cruz, Mark Grady, Ariela Gross, Jack Hirshleifer, Peter Huang, Dan Klerman, Timur Kuran, Chris Leslie, Gillian Lester, Susanne Lohmann, Tom Lyon, Paul Mahoney, Ed McCaffery, Nolan McCarty, Nancy Marder, Michael Shapiro, Dan Simon, Michael Solimine, Jeff Strnad, Cass Sunstein, Ivo Welch, and workshop participants at USC, UCLA, and Chicago-Kent law schools for helpful comments and discussions. Thanks as well to Yonit Hassid, Cindy Janis, Edgar Park, and Kathy Zeiler for excellent research assistance. All errors are mine.

1. See generally GERARD DEBREU, *THEORY OF VALUE* (1959); J. R. HICKS, *VALUE AND CAPITAL* (1946). PAUL SAMUELSON, *FOUNDATIONS OF ECONOMIC ANALYSIS* (1947); Kenneth J. Arrow & Leonid Hurwicz, *On the Stability of the Competitive Equilibrium*, 26 *ECONOMETRICA* 522 (1958).

2. Although the basic intuitions were first established in ADAM SMITH, *THE WEALTH OF NATIONS* (1776), the general-equilibrium scholarship of the 1940s and 1950s was the first to formalize it precisely.

3. This proposition has come to be known as the First Theorem of Welfare Economics. It is closely related 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).

convey valuable public information about resource scarcity to individuals possessing little more than “dispersed bits of incomplete and frequently contradictory knowledge.”⁴ By many accounts, these dual characteristics of efficiency and information transmission constitute a legacy that is central to much of modern economic thought.⁵

In light of this intellectual inheritance, 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 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,⁶ and *second* that it does so in a decidedly nonorchestrated fashion—obviating the up-front costs that constrain 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 choices. Buttressing these similarities, early advocates of the market analogy identified some striking examples within the common law in which we observe rules that seem (at least roughly) to correspond with efficiency concerns.⁷

4. F.A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519, 528 (1945).

5. Even in so-called “New Keynesian” economics, for example, 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, 1-3 (1987) (arguing that price may serve an important signal of quality).

6. See WILLIAM LANDES & RICHARD POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* (1987); RICHARD POSNER, *ECONOMIC ANALYSIS OF LAW* 320-28 (1973); William Landes & Richard Posner, *A Positive Economic Analysis of Products Liability*, 14 J. LEGAL STUD. 535 (1985); Richard Posner, *Some Uses and Abuses of Economics in Law*, 46 U. CHI. L. REV. 281 (1979) [hereinafter *Uses and Abuses*]; George Priest, *Breach and Remedy for the Tender of Nonconforming Goods Under the Uniform Commercial Code: An Economic Approach*, 91 HARV. L. REV. 960 (1978); George Priest, *The Common Law Process and the Selection of Efficient Legal Rules*, 6 J. LEGAL STUD. 65 (1977); Paul H. Rubin, *Why Is the Common Law Efficient?*, 6 J. LEGAL STUD. 51 (1977).

7. See generally Richard Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 29 (1972).

Nevertheless, among contemporary legal scholars, the analogy between market behavior and legal evolution remains relatively tenuous. Over the last twenty years, detractors of the efficiency hypothesis have challenged its foundational premises from multiple perspectives. For example, some have noted that judges (often by design) are more insulated from the inefficient consequences of their decisions than are analogous market participants. This added protection, in turn, affords them the opportunity to pursue ends that need not be related systematically to efficiency.⁸ Others have registered even stronger criticisms, noting that *even if* courts were predisposed 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.⁹

Recently, a number of scholars have posited a particularly 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.”¹⁰ 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 repeatedly ignore their own inclinations, preferring instead to emulate their predecessors. More specifically, the cascades literature posits that strategic actors may *rationally* prefer emulation, presuming (frequently incorrectly) that their own information is unreliable measured against the stock of that revealed from their predecessors’ actions.¹¹ Moreover, once an information cascade

8. See generally Robert Cooter & Lewis A. Kornhauser, *Can Litigation Improve the Law Without the Help of Judges?*, 9 J. LEGAL STUD. 139 (1980); Lewis A. Kornhauser, *Some Notes on the Logic of Legal Change*, in THE LOGIC OF SOCIAL CHANGE: ORIGINS 169 (D. Braybrooke ed., 1996); George Priest, *The Common Law Process and the Selection of Efficient Rules*, 6 J. LEGAL STUD. 65 (1977).

9. See, e.g., Gillian K. Hadfield, *Bias in the Evolution of Legal Rules*, 80 GEO. L.J. 583, 584-85 (1992).

10. See, e.g., Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683, 765 (1999) (arguing principally that public risk assessments are subject to informational, cognitive and reputational herding effects—what they define as “availability cascades”—but positing that courts are not immune to similar phenomena). See also, e.g., Andrew Daughety & Jennifer Reinganum, *Stampede to Judgment: Persuasive Influence and Herding Behavior by Courts* (Vanderbilt University School of Law working paper, 1998).

11. See, e.g., Sushil Bikhchandani, David Hirshleifer & Ivo Welch, *A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades*, 100 J. POL. ECON. 992 (1992); Abhijit V.

begins, other conformity-preserving phenomena (such as reputational concerns,¹² network externalities,¹³ status quo biases,¹⁴ informal pressures to conform,¹⁵ and social “entrepreneurs” eager to exploit group pathologies¹⁶) can amplify and entrench the initial herding effect.¹⁷ In the aggregate, 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), their 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. Beginning as early as Oliver Wendell Holmes, numerous legal scholars have maintained that the formation of precedent is symptomatic of judicial learning. The argument goes something like this: Once judges gain sufficient familiarity with a subject matter—transmitted through successive holdings in “similarly

Banerjee, *A Simple Model of Herd Behavior*, 107 Q. J. ECON. 797 (1992); Ivo Welch, *Sequential Sales, Learning, and Cascades*, 47 J. FIN. 695 (1992).

12. See generally David S. Scharfstein & Jeremy C. Stein, *Herd Behavior and Investment*, 80 AM. ECON. REV. 465 (1990); Jeff Zwiebel, *Corporate Conservatism and Relative Compensation*, 103 J. POL. ECON. 1 (1995). See also TIMUR KURAN, *PRIVATE TRUTHS, PUBLIC LIES: THE SOCIAL CONSEQUENCES OF PREFERENCE FALSIFICATION* 157-95, 289-309 (1995) (describing the interaction between information cascades, reputation, and the resulting impulse to falsify one’s own preferences).

13. See Michael Klausner, *Corporations, Contracts, and Networks of Contracts*, 81 VA. L. REV. 757, 789-840 (1995) (discussing how “network externalities” in corporate contracts can make even inefficient contractual conventions durable if a sufficiently large number of people utilize them). See also Marcel Kahan & Michael Klausner, *Path Dependence in Corporate Contracting: Increasing Returns, Herd Behavior and Cognitive Biases*, 74 WASH. U. L.Q. 347 (1996) (detailing same). But see Clayton P. Gillette, *Lock-In Effects In Law and Norms*, 78 B.U. L. REV. 813 (1998) (arguing that this “lock-in” effect understates the value of stable precedent as a mechanism for certainty).

14. See generally William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7 (1988); Russell Korobkin, *The Status Quo Bias and Contractual Default Rules*, 83 CORNELL L. REV. 608 (1998).

15. See generally George A. Akerlof, *A Theory of Social Custom, of Which Unemployment May Be One Consequence*, 94 Q. J. ECON. 749 (1980).

16. See Kuran & Sunstein, *supra* note 10, at 723-28; Donald C. Langevoort & Robert K. Rasmussen, *Skewing the Results: The Role of Lawyers in Transmitting Legal Rules*, 5 S. CAL. INTERDISC. L.J. 375 (1997).

17. Although these exacerbating effects all provide plausible rationales for the *durability* of herd-like behavior, they provide somewhat less satisfying accounts of the *genesis* of such phenomena. Information cascades, on the other hand, provide such a theory, and are therefore of central importance in assessing the *ex ante* danger posed by these alternative forms of herding behavior. I therefore concentrate principally on information cascades in this Article.

situated” cases—a binding rule may crystallize for all future cases, even if the rule is an inefficient fit for such cases. 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.¹⁸

If common law precedent is in fact a type of 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. For example, it might explain why solitary decisions such as *Lochner v. New York*¹⁹ succeeded in sowing the seeds for a string of succeeding like-minded (and non-self-conscious) opinions, only to meet their demise when their tenuous foundational underpinnings were 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

18. Oliver Wendell Holmes’ 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 Holmes did not cast his account within the rhetoric of economic efficiency, there is a fairly close parallel. See *id.* at 5 (“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. . .”). Indeed, 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 . . . enabl[es] the parties to a case, and the tribunal also, to use information that has been generated (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 that 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 according to precedent thus converge in suggesting that the practice will be consistently followed only in highly stable societies.

RICHARD POSNER, *ECONOMIC ANALYSIS OF LAW* 548 (4th ed. 1992) (emphasis added).

19. 198 U.S. 45 (1905).

Brown v. Board of Education,²⁰ as we could no longer be sure that their durability continues to reflect sound reasoning or historical path dependence.

In this Article, I endeavor to evaluate whether a cascade theory of precedent represents a cogent description of legal evolution, focusing principally on information cascades.²¹ 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 appear somewhat implausible. In particular, many long-standing institutional practices within the judiciary—such as long judicial tenures, written opinions, and the hierarchical appeals processes—tend to reduce the likelihood of a “bad” precedential cascade. Moreover, it is extremely difficult (and perhaps impossible) to verify whether observed judicial conformity is the byproduct of a cascade or of some omitted third factor that commonly affects all judges. At the same time, however, a cascade theory of precedent—even if empirically implausible—may still hold relevance for legal scholars: for not only might concerns over cascades help justify many of the existing institutional features within the judiciary, but they might also provide some limited counsel as to how judicial institutions may be designed and utilized more effectively.²² Moreover, an awareness of herding phenomena may help judges identify when a population prescribed by some legal rule may itself be prone to problematic herding behavior which a prudent judicial response might ameliorate.

I develop my argument in four parts. Part I discusses the *theoretical possibility* that precedent may constitute a cascade. By analyzing a stylized model of adjudication, in which a sequence of judges each rule on a sequence of similarly-situated cases, I demonstrate how the possibility of a “precedential cascade” might emerge if at any point in time the relevant case history becomes sufficiently monolithic. Moreover, the resulting precedent need not coincide, ironically, with the underlying normative

20. 347 U.S. 483 (1954).

21. See note 17, *supra*, for the justification of this focus. I shall nonetheless return to the alternative sources of judicial conformism at Part IV.C *infra*.

22. For example, Samuel Estreicher and John Sexton have advocated a case selection theory of certiorari for the Supreme Court that hinges on (among other things) the presence of inter-jurisdictional conflict on central federal or constitutional issues (either among federal circuits or state courts adjudicating federal claims). See SAMUEL ESTREICHER & JOHN SEXTON, REDEFINING THE SUPREME COURT’S ROLE 41-70 (1986). Conversely, Estreicher and Sexton would tend to label as “improvident” those grants of certiorari in which there was little inter-jurisdictional disagreement. See *id.* If one were mindful of the possibility of cascades, however, prudent case selection policy might cut in exactly the opposite direction. See Part IV.C *infra*.

goals of the judges creating it, thereby exposing the principal theoretical danger of information cascades. Part II discusses the more practical question of whether such herding behavior is a plausible positive account of law. Here my answer is somewhat more critical. I consider six necessary conditions 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, as a number of them are largely precluded by current institutional features of the judicial system. Part III highlights my plausibility concerns by exploring possible empirical verification techniques to determine whether legal cascades exist in at least some circumstances. My conclusions here are perhaps the most skeptical. For even if precedential cascades seemed plausible on a priori grounds, it is difficult and likely impossible to test for their existence against any number of plausible alternative hypotheses.²³ Finally, Part IV attempts to add some perspective to the foregoing arguments, discussing the viability of non-information-based accounts of precedential cascades, and arguing that however constituted, cascade theory may still shed considerable light on the institutional value of institutions such as written opinions, life tenure, the appeals process, and the like. 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?

Perhaps the first requisite step in evaluating the viability of a cascade theory of precedent is to understand precisely what an informational cascade is, and how one might emerge within a jurisprudential environment. The discussion below attempts to provide such an understanding by exploring a simple numerical model of adjudication, in which a series of like-minded judges sequentially rule on a series of similarly-situated cases.²⁴ To conform with the cascade framework, the judges are presumed rational, but incompletely informed about the world, and thus may use the holdings of their predecessors in action to inform

23. It is important to note, of course, that the current absence of empirical proof of cascades falls short of implying that cascades do not exist. It does, however, present significant problems for proponents of cascade theory if they bear the academic burden of persuasion. *See generally* THOMAS KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTION* (2d ed. 1970).

24. In 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 hinges, at least in part, on empirical learning among judges.

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, *even if* not formally required to do so, and *even if* the case they actually hear suggests a contrary outcome. Second, the incentive for judges 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 couple of 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 that (for various reasons outside the model) such precedent is binding. Rather, the ultimate question of interest here is under what circumstances will judges *rationally choose* to emulate predecessors, even if not otherwise compelled to do so. 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 (that is, those that make the analysis more tractable without altering any qualitative results) and *critical* assumptions (that is, those that do both). Indeed, a subsequent Part of this Article highlights six critical assumptions of herding models, and argues that their confluence in practice may be extremely implausible.²⁵

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,²⁶ motivated solely by economic efficiency concerns.²⁷

25. See Part II *infra*.

26. The descriptive term “Bayesian” roughly asserts that judges use available information to update and refine their beliefs about the world. In the context of this Article, judges use past and current cases to hone beliefs about the composition of the population in which 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

The tenure of each judge lasts exactly one period, at which time the judge retires and is replaced by a similarly sophisticated, rational, Bayesian, and efficiency-minded successor. Denote the series of judges as time passes by J_1, J_2, J_3 , 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.²⁸ 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. Suppose that each succeeding judge must announce a rule to govern some potentially injurious activity, whose possible harm to victims is high (for example, \$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.²⁹ The judge must decide between two potential legal rules. Under the first (the “pro-victim” rule), potential injurers are strictly liable for injuries that occur as the result of an accident.³⁰ 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 denote the rule

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:



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. And 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 an individual who has tested positive also carries the virus is equal to 20%.

27. As noted at note 24, *supra*, there is nothing critical about the assumption that judges are motivated by efficiency. So long as judges share a common motivational goal whose prescriptions turn (in part) on empirical learning, a cascade can occur—at least in theory.

28. For now, I do not attend to the question of what makes cases sufficiently similar to place them within the same “class,” assuming instead that such classifications are consistently apparent to all successive judges. When judges differ as to what constitutes a “similarly-situated case,” the dangers of herding effects significantly decrease (a point to which I shall return in Part III of the Article).

29. 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).

30. 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, and indeed some form of rule boundedness is critically important for a precedential cascade to emerge. *See infra* Part III.

announced in each period by y_i , such that a decision 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.³¹ To concentrate on this element of the judge's problem, assume that if neither party takes a precaution, then an injury-producing accident is relatively likely, at say 25%. Conversely, should *either* side exercise precaution, the likelihood of harmful accident falls very close to (but slightly above) zero.³² 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."³³

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 *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 large cost of \$6,000 to avoid harm; and "low-cost" victims, who face a relatively small avoidance cost of \$4,000. In particular, suppose that the fraction of high-cost victims within the jurisdictional population is denoted by the variable θ , which (by definition) lies somewhere between 0 and 1.³⁴ Assume (for now) that this proportion (that is, θ) remains constant over time. The respective victim types, avoidance costs, and population fractions are represented in Table 1 below:

31. As first noted by 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. See GUIDO CALABRESI, *THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS* (1970); ROBERT COOTER & THOMAS ULEN, *LAW & ECONOMICS* 347 (1988).

32. Consequently, 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 *either* of the parties takes precautions, and there is no added benefit from *both* sides taking precautions. Nothing turns on this assumption.

33. CALABRESI, *supra* note 31.

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

Type of Victim	Avoidance Cost	Fraction of Population
High-Cost	\$6,000	θ
Low-Cost	\$4,000	$(1-\theta)$

TABLE 1. VICTIM TYPES, AVOIDANCE COSTS, AND POPULATION FREQUENCY

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

$$\begin{aligned} \text{Exp. Social Costs} = & \quad \$4000 (1-\theta) + \$6000 \theta & \text{under "pro-injurer" rule} \\ & \quad \$5000 & \text{under "pro-victim" rule} \end{aligned}$$

As noted above, successive judges J_i each attempt to minimize the expression above during their tenure given their information (which is equivalent to attempting to maximize efficiency).³⁵

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 it is the former, then a pro-victim rule is optimal; if it is the latter, then the opposite holds. Moreover, heterogeneity implies that regardless of the judge's eventual opinion,³⁶ at

35. 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 borne no matter who bore the cost of precautions.

36. 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 the victim's 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 all victims are observationally equivalent to injurers, at the time of the harm the duty must (for practical reasons) extend uniformly, either to all victims or none.

least some segment of prospective injurer-victim pairs will be “mismatched” with the legal rule. Given this uncertainty, 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*.³⁷

The judge, therefore, has an incentive (using any currently-available information) to *learn* about the characteristics of the injurer-victim population. As is often typical in hard cases, however, I assume that no judge has good *ex ante* information about the true value of θ . Explicitly, everyone’s best initial guess about θ is that it could lie anywhere between 0 and 1 with uniform probability.³⁸ 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 gather more information about the relevant population characteristics, and then to prescribe a rule that best satisfies the judge’s jurisdictional preferences during her tenure.³⁹

In order to capture this intuition, suppose that at the beginning of her tenure, the judge hears exactly *one* case, drawn randomly from those in which an accident has occurred.⁴⁰ 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 employ—along with information gleaned from 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.⁴¹ Once she issues her holding, the judge’s task is complete, and

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

37. See Hadfield, *supra* note 9, for a similar approach.

38. In other words, I assume that the prior density function of θ is uniform on the unit interval.

39. Note that if θ 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 whose beliefs about θ make her indifferent will simply randomize between the two rules.

40. 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 θ are of type 1. In Part II, I consider the effects of allowing judges to hear more than one case, which affects the outcome.

41. Because the judge is assumed motivated solely by efficiency concerns (rather than compensatory goals), the fact that the harm has already occurred implies that her efficiency considerations will be exclusively prospective in nature. A similar story could be told, however, if the judge was motivated by distributional concerns *if* she were attempting to learn information about wealth distributions among injurer-victim pairs.

her announced rule governs the jurisdiction until the end of the period, upon which the judge is replaced with a successor.

As noted above,⁴² one interesting consequence of viewing adjudication as a source of judicial 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.⁴³

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 proscribed population of actors. As will become apparent, precedent (at least within this example) 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 established 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 substantial prior information about the characteristics of the victim population (that is, θ).⁴⁴ With such limited information, other than the facts she observes in the first case, the judge has very little on which to depend. Thus the facts of the instant case may play a pivotal role in determining her ultimate decision. Indeed, it is straightforward to demonstrate⁴⁵ 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

42. See *supra* notes 38-40 and accompanying text.

43. This “predictive” element of precedent is consistent with the description often ascribed to legal realists. See RONALD DWORKIN, *LAW’S EMPIRE* 153 (1986).

44. Formally, it is necessary to specify a status quo ante that governs individual behavior even prior to the first judge’s opinion. I shall assume (somewhat arbitrarily) that this status quo ante corresponds to a “state of nature” in which injurers are never liable for harms they cause. Nothing turns on this assumption.

45. See Appendix *infra*.

holding is somewhat informative to her successors: for it signals precisely the type of victim she has observed.

Let us now assume, for argument's sake, that the second-period judge, J_2 , inherits a history in which the first judge has held for the pro-victim rule (thus indicating that J_1 observed a high-cost victim). If J_2 also observes 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, J_2 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 valuable to the second judge. However, they jointly constitute evidence 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 the pro-victim and the pro-injurer rule.⁴⁶ 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 still 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 (that is, $y_1=y_2=1$). What might she infer from this history? Certainly, she can infer, as did J_2 , that the first judge 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, J_2 , *probably* also observed a high-cost victim.⁴⁷ This inference, as it turns out, is a strong one. It is sufficiently 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

46. This assumption is conventional in game theory, and can easily be relaxed.

47. Given what J_1 reveals about θ , though, it is possible that J_2 observed a low-cost victim case and randomized. However, given that J_2 certainly observed a high-cost victim, this interpretation is less likely than the alternative that J_2 likewise observed another high-cost victim.

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 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 lopsided in favor of either the pro-injurer or pro-victim rules, a precedential cascade will begin and will continue indefinitely.⁴⁸

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, once 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%, and thus the majority of victims (that is, 60% 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, there is no guarantee that a stable precedent (when it obtains) will coincide with the most efficient rule. Indeed, it is possible to specify a number of situations in which a cascade emerges around the less efficient doctrine. For example, should the first two judges both hold in favor of the relatively inefficient pro-victim rule (a

48. In particular, in this framework, a cascade will begin as soon as the number of pro-victim holdings outnumbers the number of pro-injurer holdings by two (and vice-versa). See Appendix *infra*. Moreover, it can be shown that in this sort of framework, a precedential cascade eventually emerges with probability 1. See *id.*

contingency that occurs 28% of the time⁴⁹), a cascade will begin immediately around that rule.⁵⁰

Figure 1 below illustrates the probability (measured from the ex ante perspective) that the jurisdiction's prevailing rule will be the inefficient one during each successive period. Note that during the initial periods, the probability of an inefficient rule rapidly falls. But this seemingly convergent process quickly slows in subsequent periods as the likelihood that a cascade has already begun increases. In fact, the probability that the jurisdiction adopts an inefficient rule never falls below 36.84%, *no matter how many periods one considers*.

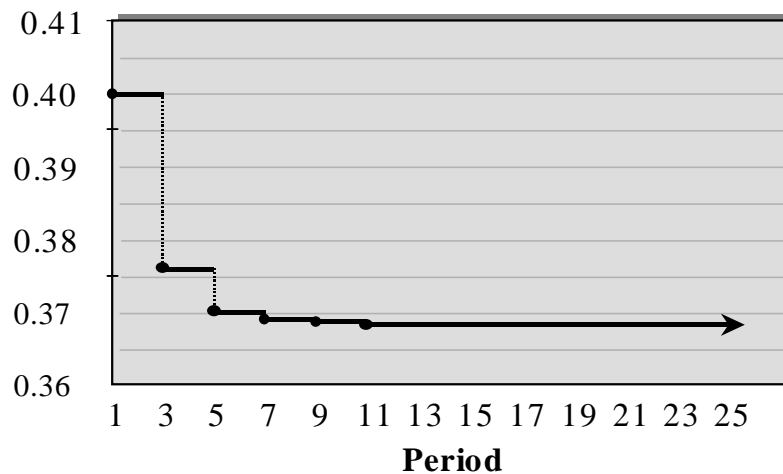


FIGURE 1: EX ANTE PROBABILITY OF AN INEFFICIENT RULE EACH PERIOD

A few characteristics from this example bear emphasizing. First, 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, knowing that her predecessor has (rationally) decided to emulate her predecessors, each succeeding judge finds herself in essentially the same position as her predecessor, and would therefore also find it optimal to

49. Explicitly, such a history would occur whenever the first judge sees a high-cost victim (which happens with probability 0.4) and either (1) the second judge sees the same (probability 0.4); or (2) the second judge observes a low-cost victim, randomized between the two rules, and ends up selecting the pro-victim rule (probability $0.6 \times 0.5 = 0.3$). The joint probability of the sequence, then, is equal to $(0.4) \times (0.4 + 0.3) = 0.28$, or 28%.

50. Moreover, as noted above, if at *any* time after the first two periods pro-victim holdings become sufficiently numerous, an inefficient cascade will occur at that time. See *supra* note 48.

follow suit. In turn, so must all judges 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—for argument’s sake—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 J_3 actually observed 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 in a cascade 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 decisionmakers, 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).⁵¹

Second, note that despite its thin empirical basis, a stable precedential cascade tends to emerge rather rapidly in this example. Indeed, in the example depicted in Figure 1, there is approximately a 76% chance that a stable precedent will have emerged by the third period, growing to 98% by the seventh period. Nevertheless, whenever such a cascade starts, *it will be the wrong one nearly two-fifths of the time*. Consequently, a rapid convergence to a stable precedent reveals little about whether 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 an initial 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,”⁵² or rather because of a chance dependence on initial cases that turn out to be statistical outliers.⁵³

Third, although an information cascade (once commenced) appears stable, it is also rather brittle. Information cascades are extremely path-dependent on the outcomes of the first few cases. Consequently, if an external shock from outside the above framework (caused by a public disclosure of information, drift in population demographics, or the like)

51. Some commentators seem to overlook this point. See, e.g., Kuran & Sunstein, *supra* note 10, 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.”).

52. See *Uses and Abuses*, *supra* note 6.

53. A number of legal scholars have argued that the *lack* of path-dependence is a desideratum of a common-law system. See, e.g., FRANK H. EASTERBROOK & DANIEL R. FISCHER, *THE ECONOMIC STRUCTURE OF CORPORATE LAW* (1991).

were to occur, the cascade might easily “break,” as its thin empirical basis is finally exposed.⁵⁴ I take these issues up more explicitly in Part IV, *infra*.

Finally, while the example above illustrates how precedents may diverge from *efficiency* goals, the implications of the herding approach—at least in theory—need not be limited to a single motivational account of judging. Indeed, suppose, for example, that judges were uninterested in efficiency, but rather 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.⁵⁵ 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 (that is, 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

54. On the other hand, it is important to keep in mind that informational cascades do not occur in a vacuum. Information cascades might well trigger other behavioral and cognitive pathologies, which could reinforce conformism. Of particular note here is the “availability heuristic,” which describes the proclivity of individual decisionmakers to bias excessively their probabilistic assessments of some random phenomenon in accordance with their ability to recall specific instances of that phenomenon occurring. See generally Kuran & Sunstein, *supra* note 10. See also TIMUR KURAN, PRIVATE TRUTHS, PUBLIC LIES: THE SOCIAL CONSEQUENCES OF PREFERENCE FALSIFICATION 157-75 (1995) (elaborating on this concept in greater detail); ANTHONY PRATKANIS & ELLIOT ARONSON, AGE OF PROPAGANDA: THE EVERYDAY USE AND ABUSE OF PERSUASION 134-39 (1992) (examining the use of the same advertisements repetitively); Lynn Hasher, David Goldstein & Thomas Toppino, *Frequency and the Conference of Referential Validity*, 16 J. VERBAL LEARNING & VERBAL BEHAV. 107, 107-12 (1977); Marian Schwartz, *Repetition and Rated Truth Value of Statements*, 95 AM. J. PSYCH. 393, 393-407 (1982).

55. There is, of course, a lively debate (even within law and economics) about whether courts should pursue anything other than efficiency goals. Compare Louis Kaplow & Steven Shavell, *Why the Legal System Is Less Efficient Than the Income Tax in Redistributing Income*, 23 J. LEGAL STUD. 667 (1994) (arguing that the pursuit of distributional goals should be the province of the tax system alone), with Chris William Sanchirico, *Taxes versus Legal Rules as Instruments for Equity: A More Equitable View* (USC Law School Working Paper No. 98-21, 1998) (visited October 2, 1999) <http://papers.ssrn.com/paper.taf?ABSTRACT_ID=47400> (arguing that Kaplow & Shavell’s argument depends on an unrealistic critical assumption). This debate, however, is not one that is material to this Article.

(efficiency-oriented or otherwise) that envisions some element of judicial learning.⁵⁶

II. ARE PRECEDENTIAL CASCADES PLAUSIBLE?

To be sure, the theoretical 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, as described above, are a plausible description of the judicial process. I now attend to this matter, focusing on six necessary conditions for “bad” precedential cascades to occur⁵⁷: rule-boundedness, decisional opacity, judicial homogeneity, short judicial tenures, flat hierarchies, and population stationarity. Not only is each of these conditions far from ubiquitous when viewed in isolation, but their mutual confluence is likely to be *especially* rare. Moreover, a number of these conditions are avoidable by the prudent design of a judicial system, manifestations of which are already largely present within the current institutional regime.⁵⁸

A. RULE-BOUNDEDNESS

For a precedential cascade to occur, it is necessary that judges be “rule-bound.” In other words, judges must have a limited number of legal rules to choose from. 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.

56. See generally, e.g., FREDERICK SCHAUER, *PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE BASED DECISIONMAKING IN LAW AND IN LIFE* (1991).

57. In the formal sense of the word, some of the conditions described below are not strictly “necessary,” as a cascade might still occur in their absence. Nevertheless, I include them below because their absence would still render a cascade phenomenon particularly unlikely.

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

As a result, the information transmitted by a rule-bound judge to her successors contains so much noise as to have no informational value.⁵⁹

Undoubtedly, rule-bound judging may be an accurate description of some legal doctrines.⁶⁰ Liability, for example, is often a binary outcome by definition.⁶¹ 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 example⁶²), 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. Courts must also make determinations as to when the “facts” of a case are sufficiently similar to those that the applicable rule is supposed to govern.⁶³ Each of these ancillary decisions may represent a valuable source of

59. In 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 would 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). One small caveat is applicable here: Cascades can conceivably occur when the judges' choice set is continuous, but bounded, such as the real number line between 0 and 1. If the extremes of such a choice set (i.e., 0 and 1) represent policy choices that judges might plausibly choose in some certain states of the world, then there is a small probability that a cascade will occur at one of those endpoints. See Daughety & Reinganum, *supra* note 10. Nevertheless, such an environment still implies that judges are effectively rule-bound at the boundaries of their action space.

60. Many pragmatists, such as Cass Sunstein, in fact, favor casuistry over rule-boundedness. See CASS SUNSTEIN, *LEGAL REASONING AND POLITICAL CONFLICT* 136-47, 191-96 (1995). On the other hand, Sunstein admits the necessity of rules in many cases, including default (and other “privately adaptable”) rules within contracts and corporations. See *id.* at 154-58.

61. In 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 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 straightforward 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 generally EASTERBROOK & FISCHER, *supra* note 53.

62. For 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).

63. Recall that in the example from Part II this classificatory decision was essentially assumed away. See *supra* note 28.

information to later judges, even if the instant judge is unwilling to dislodge the liability rule. When incorporated into the example from Part I, such arguments can 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.

B. DECISIONAL OPACITY

Another necessary condition for cascades is that the communication between actors be limited. For example, in stock market applications of herding behavior, market participants are able to observe transactions only, not the motivating rationales of the buyers and sellers.⁶⁴ Similarly, a precedential cascade implicitly assumes that succeeding judges are able to observe only what rule predecessor courts have adopted—but not *why* they have adopted it.⁶⁵ On its face, such an assumption within a judicial system is *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 written opinions 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 a reluctance to do 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 victim (rather than the injurer) is the low-cost risk avoider. Clearly, this single observation may not be enough to induce J_3 to break with precedent, but a written opinion would allow her both to follow her predecessors' lead *and* to signal her ambivalence about doing so.⁶⁶ 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 after the third period.

64. See Lee, *supra* note 59, at 410.

65. See, e.g., Banerjee, *supra* note 11, at 798.

66. See, e.g., *Lake River Corp. v. Carborundum Co.*, 769 F.2d 1284, 1289 (7th Cir. 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").

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 illustrates, when written opinions are available, judicial learning will not stagnate as it did earlier. Indeed, if one were to let this process continue indefinitely, the prevailing legal rule will converge (with a probability of one) to the most efficient legal rule.⁶⁷

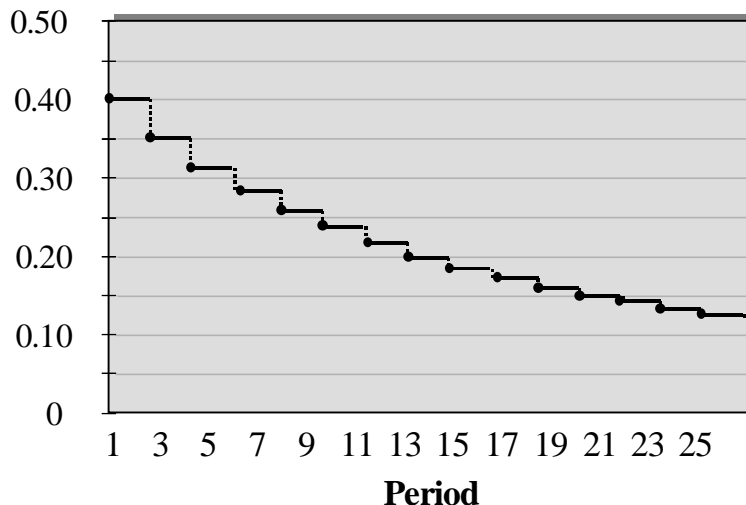


FIGURE 2: EX ANTE PROBABILITY OF INEFFICIENCY (WRITTEN OPINIONS)

It is, of course, debatable whether written opinions are always perfectly candid reflections of a judge's thoughts. Indeed, 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.⁶⁸

67. One caveat to this argument deserves mention. In the example from the previous section, it was assumed that the judge had no choice but to observe the facts of the case before her. When, however, such observations are at the discretion of the judge (e.g., she can dismiss the case at a relatively early stage), then a judge caught in a cascade may rationally decline even to collect the instant signal. See L. Smith & P. Sorenson, *Pathological Outcomes of Observational Learning* (MIT Working Paper, 1995) (on file with author).

68. See, e.g., THURMAN ARNOLD, *THE SYMBOLS OF GOVERNMENT* (1935); CHARLES CLARK, *REAL COVENANTS AND OTHER INTERESTS WHICH RUN WITH THE LAND* (2d 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 COLUM. L. REV. 809, 812 (1935) (making a similar argument about personal jurisdiction doctrine); Max Radin, *The Theory of Judicial Decision: Or How Judges Think*, 11 A.B.A. J. 357, 360 (1925). Karl Llewellyn's now-classic article on judicial interpretation of statutes underscores this point, as Llewellyn notes some 28 separate pairs of "canons of construction," each pair seeming to suggest opposite behavioral norms, such as the

Moreover, recent “pragmatist” theory has even endorsed the idea of opaque decisions. Cass Sunstein, for example, has argued (somewhat ironically) that judges *should* be somewhat imprecise, confining the justificatory components of their holdings to “low-level” principles that are consistent with numerous normative theories of law.⁶⁹ 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.⁷⁰ 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

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.” Karl Llewellyn, *Remarks on the Theory of Appellate Decision and the Rules or Canons About How Statutes Are to Be Construed*, 3 VAND. L. REV. 401, 404 (1950).

69. See CASS SUNSTEIN, *LEGAL REASONING AND POLITICAL CONFLICT* 193-94 (1995). 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 of “bad” cascades. See Kuran & Sunstein, *supra* note 10, at 762 (“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).

70. 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, expose the normative commitments 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 fact that is critical to efficiency-minded concerns), and who, between the injurer and the victim, is the wealthiest (a fact that is critical to distributional concerns). Sunstein’s account, as I read it, suggests that this judge may prefer to issue a relatively uniformative holding (e.g., “the injurer is liable because consideration of justice tip in favor of liability”) 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 implicit recitations of facts, this judge may be able to garner support from judges who pursue distinct (but in this case consequentially equivalent) norms.

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.

C. JUDICIAL HOMOGENEITY

A third necessary condition for a cascade is that the motivational objectives of all actors must be mutually shared and commonly known. In the model from Part I, for instance, it was common knowledge that all judges were identical, one-period efficiency maximizers, their only difference being the order in which they decide cases. This motivational homogeneity enabled later judges to make inferences about their predecessors' observations 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 was 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 the case before her, thereby breaking the possible cascade.⁷¹ This dampening effect is even stronger when judges are unsure of each other's motivations. Asymmetric information of this sort magnifies the

71. Kuran & Sunstein, *supra* note 10, 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"). To be sure, individuals must act sequentially (rather than simultaneously) for a cascade to form, but heterogeneity beyond sequential order of action is more likely to dampen (rather than amplify) herd-like behavior. For instance, if individual judges were heterogeneous in their proclivity to join the bandwagon, a cascade would tend to build slowly, as less responsive individuals slowly join the bandwagon. However, were an apparent bandwagon effect to erupt in this fashion, it would be less likely to constitute a problematic cascade. Indeed, less responsive individuals (by definition) have a smaller proclivity to infer wisdom from their predecessors in action, and pay more attention to their own private signals. In this sense, the fact that a relatively unresponsive individual joined the herd late signals valuable new information to successors (unlike the stagnated learning that typifies information cascades).

Nevertheless, heterogeneity 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. *See* Zwiebel, *supra* note 12. Note, however, that even within this framework, heterogeneity makes population-wide conventions *less* rather than more likely. *See infra* Part IV.C.

incentive to place one's own observation above the (imperfectly revealed) observations of others.⁷²

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.”⁷³ 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 decisionmakers seek to maximize their own benefit by following the herd, disregarding the welfare of future decisionmakers, who would benefit from learning about predecessors' observations.

One oft-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.⁷⁴ 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.⁷⁵

As it turns out, common law judicial systems have an excellent device for ensuring repeat play: long judicial tenures.⁷⁶ 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,

72. 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 *dynamic* rather than *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.

73. See generally Garrett Hardin, *The Tragedy of the Commons*, 162 SCI. 1243 (1968).

74. See *id.* at 1245.

75. Of course, it may still be possible for an infinitely-lived judge to bias her recollection of previous cases in the direction of particularly note-worthy experiences. See *infra* Part IV.C.

76. Life tenure has a longstanding tradition within Anglo-American jurisprudence. See generally Daniel Klerman, *Nonpromotion and Judicial Independence*, 72 S. CAL. L. REV. 455 (1999).

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% to 78%, and the corresponding probability that an inefficient precedent would emerge decreases from 37% to 22%.⁷⁷ Thus, a theory of precedential cascade is far less compelling when individual judges serve out long terms on the bench.

E. FLAT HIERARCHIES

Another important assumption made in the previous Section was that the hierarchy of courts was relatively flat. In particular, within that example there was only one court in each period, whose holding was not subject to subsequent review by a higher court. Quite clearly, such flat hierarchies are much more the exception than the rule within Anglo-American jurisprudence. If one were to add a hierarchical judicial system to the example from Part I, the dangers presented by cascade-like behavior would almost certainly decline, for at least three reasons. First, the appeals process itself is an institutional device that facilitates the pooling of information, particularly when—as is not uncommon—appellate courts hear numerous appeals simultaneously. By reviewing the composite records of such cases, an appellate judge may be able to extract a more informative signal than did lower court counterparts, breaking a destructive

77. 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 force 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 noninvestigatory capacity of the court suggests 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.,* *United States v. Topco Assocs.*, 405 U.S. 596, 611-12 (Brennan, J., concurring) (“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 under 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 ALLAN WRIGHT, ARTHUR R. MILLER & EDWARD H. COOPER, *FEDERAL PRACTICE AND PROCEDURE* §§ 3532-34 (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. *See* FED. R. EVID. 401.

information cascade if such an action is warranted.⁷⁸ Second, appellate-level judges may be more experienced, more skilled, and less time-constrained on average relative to their trial-court counterparts. As such, they may be able to filter out some of the observational noise that a trial court would not be able to eliminate. Finally, the prospective threat of reversal may induce lower courts to draft more comprehensive and precise opinions, which in turn provides appellate-level judges with a more complete record from which to synthesize such empirical information. Viewed in this sense, then, the hierarchical structure of the judicial system places an important check on the viability of runaway herding effects. Although stacked hierarchies cannot completely circumvent cascade, such structures almost certainly render them less likely to occur.

F. POPULATION STATIONARITY

Finally, for an information cascade to persist over time, there must not be a significant element of “drift” in the underlying population characteristics of interest. In the example from Part I, for instance, the absence of drift implies that the population of injurers and victims—and in particular the ratio of high-cost to low-cost victims—must remain relatively constant over time. If it did not, then it would be difficult for late-moving judges to learn much, if anything, about early judicial opinions. Indeed, such opinions might have been statistically justifiable at the time, but 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, episodic cascades could emerge. However, such cascades would be relatively less durable, 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 precedential cascades less troublesome.

78. Indeed, 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. See Daughety and Reinganum, *supra* note 10, at 4 *passim*.

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 subuniverse 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 (or unavailable), *and* population drift is relatively small. Perhaps such a subuniverse 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 informational herding is implausible as a general account of judicial review, its contributions might still have considerable value if appropriately limited to those situations where the dangers of such phenomena appear significant. Indeed, if it were possible to isolate a strand of legal doctrine swept up 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).⁷⁹ In substantive areas of 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.

The practical implementability of such 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 cross-sectional and/or time-series data on the formation of precedent 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 phenomenon. Unfortunately, neither of these approaches clearly demonstrates that an episode of apparent group conformity stems from a cascade or some other, less distressing explanation.

79. See, e.g., Kuran & Sunstein, *supra* note 10, at 688.

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,⁸⁰ corporate conservatism,⁸¹ bank runs, and political revolutions.⁸² 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 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 does not lack for rapidly-spreading doctrines. A particularly good example of such a trend is the spread of the contract doctrine of unconscionability during the last forty years.⁸³ The doctrine, in its modern form at least, allows a party to void a contractual term(s) 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).⁸⁴ 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 1960.⁸⁵ The doctrine soon began to spread to other jurisdictions, many of which cited the early movers in almost a cursory fashion.

80. See, e.g., Jeremy Bulow & Paul Klemperer, *Rational Frenzies and Crashes*, 102 J. POL. ECON. 1 (1994).

81. See, e.g., Zwiebel, *supra* note 12.

82. See Susanne Lohmann, *The Dynamics of Informational Cascades*, 47 WORLD POL. 42 (1994).

83. The diffusion of legal doctrines across states has itself spawned an entire literature that is distinct from the informational herding literature. See, e.g., Bradley C. Canon & Lawrence Baum, *Patterns of Adoption of Tort Law Innovations: An Application of Diffusion Theory to Judicial Doctrines*, 75 AM. POL. SCI. REV. 975 (1981) (analyzing the diffusion of various tort doctrines across states).

84. Most opinions that invoke the doctrine involve facts that have both procedural and substantive elements, and a strong showing of one tends to permit a weaker showing of the other. See E. ALLEN FARNSWORTH, *CONTRACTS* 312 (3d ed. 1999). It is somewhat doubtful, however, that one may utilize the doctrine in situations in which only one form of unconscionability is present. See *id.*

85. See *Henningsen v. Bloomfield Motors, Inc.*, 161 A.2d 69 (N.J. 1960). One of the first articulations of the doctrine is in the Uniform Commercial Code § 2-302, drafted during the 1950s. Though the U.C.C. constitutes compelling authority only for cases of goods, the adoption of the doctrine described here was much broader than that, and applied U.C.C. § 2-302 by analogy to other cases as well.

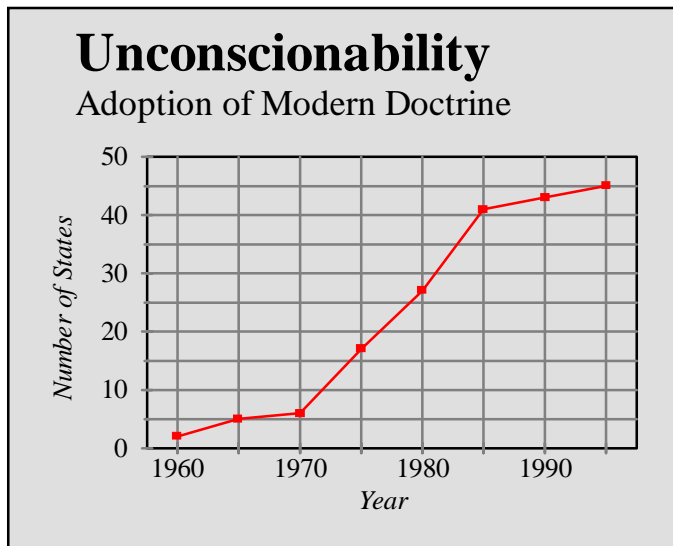


FIGURE 3: THE SPREAD OF THE UNCONSCIONABILITY DOCTRINE

The swiftness of the unconscionability trend was relatively dramatic, though not simultaneously adopted, as depicted in Figure 3. In 1970, the doctrine had been adopted by at most six states. By 1985, however, forty-one states had embraced some version of the doctrine completely. Since then, the trend has obviously slowed, and by 1996 only five more states had completely embraced the doctrine.⁸⁶ Like other purported episodes of

86. Though the term “unconscionable” has made early and frequent appearances in contract law cases, I confined my analysis to the adoption of the “modern” doctrine, in which courts recognized unconscionability formally (combining procedural and substantive elements). The state-by-state breakdown as of 1996 is follows: Ala.: *Lloyd v. Service Corp. of Ala. Inc.*, 453 So. 2d 735 (1984); Alaska: *Inman v. Clyde Hall Drilling Co.*, 369 P.2d 498 (1962); Ariz.: *Seekings v. Jimmy GMC of Tucson, Inc.*, 638 P.2d 210 (1981); Ark.: *Kohlenberger, Inc. v. Tyson’s Foods Inc.*, 510 S.W.2d 555 (1974); Cal.: *Graham v. Scissor-Tail, Inc.*, 623 P.2d 165 (1981); Colo.: *Jones v. Dressel*, 623 P.2d 370 (1981); Conn.: *Hamm v. Taylor*, 429 A.2d 946 (1980); Del.: *Tulowitzki v. Atlantic Richfield Co.*, 396 A.2d 956 (1978); Fla.: Yet to adopt; Ga.: *R.L. Kimsey Cotton Co., Inc. v. Ferguson*, 214 S.E.2d 360 (1975); Haw.: *Earl M. Jorgensen Co. v. Mark Construction, Inc.*, 540 P.2d 978 (1975); Idaho: *Brooks v. Terteling*, 688 P.2d 1167 (1984); Ill.: *Streams Sports Club, LTD v. Richmond*, 457 N.E.2d 1226 (1983); Ind.: *Weaver v. American Oil Co.*, 276 N.E.2d 144 (1971); Iowa: *C. & J. Fertilizer, Inc. v. Allied Mut. Ins. Co.*, 227 N.W.2d 169 (1975); Kan.: *Steele v. J.I. Case Co.*, 419 P.2d 902 (1966); Ky.: Yet to adopt; La.: Yet to adopt; Me.: *Dairy Farm Leasing Co., v. Hartely*, 395 A.2d 1135 (1978); Md.: *Williams v. Williams*, 508 A.2d 985 (1986); Mass.: *Zapatha v. Dairy Mart, Inc.*, 408 N.E.2d 1370 (1980); Mich.: Yet to adopt; Minn.: Yet to adopt; Miss.: *Johnson v. Robinson*, 351 So. 2d 1339 (1977); Mo.: *Bracey v. Monsanto Company, Inc.*, 823 S.W.2d 946 (1992); Mont.: *All-States Leasing Co. v. Top Hat Lounge, Inc.*, 649 P.2d 1250 (1982); Neb.: *Melcher v. Boesch Motor Co.*, 198 N.W.2d 57 (1972); Nev.: *Bill Stremmel Motors, Inc. v. IDS Leasing Corp.*, 514 P.2d 654 (1973); N.H.: *American Home Improvement v. MacIver*, 201 A.2d 886 (1964); N.J.: *Henningsen v. Bloomfield Motors*, 161 A.2d 69 (1960); N.M.: *Guthmann v. La Vida Llana*, 709 P.2d 675 (1985); N.Y.: *Equitable Lumber Corp. v. IPA*

herding, the movement toward unconscionability exhibited some weak geographic diffusion characteristics, beginning in the Northeastern and New England states, and spreading to other states later.⁸⁷

Does this evidence suggest that the doctrine of unconscionability is a form of information cascade? It certainly *may*. Indeed, a number of early decisions appear to cite other states' cases liberally,⁸⁸ indicating at least some informational interdependence between jurisdictions. At the same

Land Dev. Corp., 344 N.E.2d 391 (1976); N.C.: Brenner v. Little Red Sch. House LTD, 274 S.E.2d 206 (1981); N.D.: Haugen v. Ford Motor Corp., 219 N.W.2d 462 (1974); Ohio: Lake Ridge Academy v. Carney, 613 N.E.2d 183 (1993); Okla.: B.E. Barnes v. Helfenbein, 548 P.2d 1014 (1976); Or.: W.L. May Co. v. Philco-Ford Corp., 543 P.2d 283 (1975); Pa.: Witmer v. Exxon Corp., 434 A.2d 1222 (1981); R.I.: Grady v. Grady, 504 A.2d 444 (1986); S.C.: Yet to adopt; S.D.: Johnson v. John Deere Co., 306 N.W.2d 231 (1981); Tenn.: Ford Motor Co. v. Moulton, 511 S.W.2d 690 (1974); Tex.: Chastain v. Koonce, 700 S.W.2d 579 (1985); Utah: Bekins Bar V Ranch v. Huth, 664 P.2d 455 (1983); Vt.: Lamoille Grain Co. v. St. Johnsbury, 369 A.2d 1389 (1976); Va.: Yet to adopt; Wash.: Schroeder v. Fageol Motors, 544 P.2d 20 (1975); D.C.: Williams v. Walker-Thomas Co., 350 F.2d 445 (D.C. Cir. 1965); W. Va.: Ashland Oil, Inc. v. Donahue, 223 S.E.2d 433 (1976); Wis.: Discount Fabric House v. Wisconsin Tel. Co., 345 N.W.2d 417 (1984); Wyo.: *In re Estate of Frederick*, 599 P.2d 550 (1979).

87. The adoption of unconscionability does not seem to track closely with standard demographic predictors outside of geography. A logistic regression of demographic variables of pre-1986 adoption does not reveal a strong nongeographic element:

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

* $p < 0.10$; ** $p < 0.05$

N = 51 (50 states & District of Columbia)

Source for Demographic data: U.S. Statistical Abstract (1995).

The only independent variable showing statistical significance is membership in southern states, it being weakly predictive of a failure to adopt. The qualitative results reported above persist with a number of other permutations in the model (in which one or more of the independent variables are excluded). Moreover, the demographic variables in the specification above do not have a statistically significant *joint* effect on pre-1986 adoption. $\chi^2(46) = 47.346$; $p=0.417$. The apparent lack of strong demographically-oriented diffusion as an empirical matter accords with similar findings with regard to other doctrines. *See, e.g.*, Canon & Braum, *supra* note 83, at 983-84 (noting similar evidence for geographic diffusion in tort law). *But see* James M. Lutz, *Regional Leaders in the Diffusion of Tort Innovations Among the American States*, 27 PUBLIUS 39 (1997) (arguing that there is some evidence of required leadership in judicial innovation).

88. Of the adopting states listed in note 86 *supra*, over three-quarters cite to either *Henningsen v. Bloomfield Motors* or *Williams v. Walker-Thomas Co.*, as persuasive authority.

time, however, one of the largest difficulties one faces with rational herding models is that their causal triggers are difficult to verify empirically in the face of other plausible hypotheses. For example, it may be that a detected trend toward conformity is an artifact of some population-wide informational “shock” eliciting rapid cross-sectional adjustment, but one that outsiders might fail to measure or detect. Indeed, the timing of many of the “watershed” unconscionability cases (such as *Williams v. Walker-Thomas Furniture*⁸⁹) coincided with the publication of numerous high-profile empirical studies of poverty in the United States—tending to suggest that the relative plight of the American poor was worse than many had thought.⁹⁰ The timing also coincided with the arrival of a number of Left-Democrat-appointed 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.⁹¹ These alternative phenomena are neither byproducts nor causes of cascades, but either could just as easily give rise to the trend toward conformity noted above.

This statistical identification problem is significant, particularly for those interested in normative legal reform. Indeed, it may be 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 probably insufficient. Moreover, while the discussion above is limited to a particular doctrinal candidate for precedential cascades (that is, the unconscionability doctrine), the verification problem is more general: Because herding phenomena are the handiwork of information asymmetries that cannot be directly observed by a statistician, 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.⁹²

89. 350 F.2d 445 (D.C. Cir. 1965).

90. Of particular importance was the work of Michael Harrington. See generally MICHEAL HARRINGTON, *THE OTHER AMERICA* (1962).

91. Indeed, Skelly Wright, the author of *Williams v. Walker-Thomas*, was himself a Democratic appointee (Truman) to the federal bench, having been elevated to the D.C. circuit by President Kennedy in 1962.

92. Matters may be different with experimental data, however, where the environmental factors at play may be controlled.

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 on determining whether a series of judicial opinions constituting a possible cascade reflects (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 champions 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*⁹³ might have constituted a precedential cascade.⁹⁴ 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.⁹⁵ 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.”⁹⁶

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 predecessors for support. However, none are lacking in long and even protracted analysis. (The First Circuit’s opinion in *Eastern Enterprises*, for example, is nearly thirteen two-column pages

93. 524 U.S. 498 (1998).

94. See Daughety & Reinganum, *supra* note 10, at 5-7.

95. See *Eastern Enters. v. Chater*, 110 F.3d 150 (1st Cir. 1997); *Blue Diamond Coal Co. v. Shalala*, 79 F.3d 516 (6th Cir. 1996); *Davon, Inc. v. Shalala*, 75 F.3d 1114 (7th Cir. 1996); *Holland v. Keenan Trucking Co.*, 102 F.3d 736 (4th Cir. 1996); *Lindsey Coal Mining Co. v. Chater*, 90 F.3d 688 (3d Cir. 1996); *Barrick Gold Exploration, Inc. v. Hudson*, 47 F.3d 832 (6th Cir. 1995); *Chateaugay Corp. v. Shalala*, 53 F.3d 478 (2d Cir. 1995); *Unity Real Estate Co. v. Hudson*, 977 F. Supp. 717 (W.D. Pa. 1997); *Unity Real Estate Co. v. Hudson*, 889 F. Supp. 818 (W.D. Pa. 1995).

96. *Eastern Enters.*, 110 F.3d at 152.

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*.⁹⁷ As noted above, a precondition for a cascade is that such facts cannot be communicated among judges outside of their rule-bound holdings. As to the second question, 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.⁹⁸ 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. Such 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 “right” normative trajectory. However, in the absence of a reliable mechanism for diagnosing and verifying whether such strands of case law constitute bad precedential cascades, it is perhaps most prudent to approach legal reform proposals purporting to ameliorate such problems with significant caution. Failure to do so may give rise to even more severe lapses of judgment.

IV. WHAT REMAINS?

The previous Sections of this Article have argued that although precedential cascades are theoretically possible, their occurrence seems implausible under current judicial practices and probably unverifiable with observable data. Notably, my analysis has thus far avoided the question of whether—in light of these arguments—a cascade theory of precedent contributes anything to legal theory. It is to that question that I now attend. Although the criticisms articulated above are strong ones, it does *not* necessarily follow that the cascades literature is irrelevant for legal theory, at least within a properly-specified domain. In particular, there are at least

97. See, e.g., *id.* at 161 n.8 (“Finally, *Unity Real Estate*, on its facts, see 889 F. Supp. at 829-31, is easily distinguishable from the case at bar.”).

98. See *Eastern Enters.*, 524 U.S. at 498.

three distinct ways in which cascades can be of tremendous assistance to legal scholars, legal reformers, and judges. First, a cascade theory of precedent (as described above) may provide a positive account of many existing legal institutions as cascade-preventing devices. Second, an awareness of the potential dangers of cascades can provide an important (though perhaps limited) normative compass, both in evaluating legal reform proposals, and in understanding group conformity among populations that the law regulates. And finally, the criticisms raised above help locate and sharpen the diagnostic debate over the dangers of judicial conformity: for if the likelihood of *information* cascades among judges is small, then the most serious dangers of conformity (if any) must emanate from other, non-information-based accounts of herd-like behavior. I shall briefly address each in turn.

A. CASCADES AND POSITIVE ANALYSIS

Even if the incidence of a precedential cascade seems unlikely given current judicial practices, its theoretical possibility may nonetheless hold important implications for positive theories of law. In particular, the lurking specter of information cascades may help us to better understand and appreciate existing judicial institutions that work to attenuate such dangers. Institutional practices that may appear wasteful or redundant to the untrained eye may in actuality serve as valuable anti-cascade devices. As noted in Part II, the practice of written opinions, the hierarchical appeals process, the length of judicial tenures, the heterogeneity of the judiciary (which long tenure facilitates), and the availability of standards and fine-tuned rules all tend to minimize the likelihood of a precedential cascade. And while such practices probably did not come about *because* of concerns over cascades, the durability of these institutions may well be an indication of their effectiveness in preventing herding.

Consider, for example, the judicial practice of justifying decisions with written opinions, which began in the United States a little over two centuries ago.⁹⁹ The anti-herding effects of the written opinion requirement are powerful ones, consisting of two separate components.

99. In 1785, Connecticut became the first state to require written opinions by judges. See ERWIN C. SURRENCY, *A HISTORY OF AMERICAN LAW PUBLISHING* 42 (1990); CHARLES WARREN, *A HISTORY OF THE AMERICAN BAR* 328 (1911). Although the practice eventually became universal, the diffusion of the written opinion requirement was relatively slow (an interesting non-characteristic of a cascade). Indeed, by 1815, only six states published official court reporters, limiting their publication only to appellate courts. Towards the end of the 19th Century, however, written opinions had become the norm. See Surrency, *supra* at 42-43.

First, the writing requirement effectively forces a judge first to collect and synthesize the pertinent facts of the case before her. If not otherwise required to do so, a judge caught within a cascade would no doubt realize that her decision in the instant case would not turn on any observations about that case, and she might therefore have little incentive to collect and process them.¹⁰⁰ The written opinion norm helps counteract this adverse incentive, effectively making collection of information compatible with the judge's incentives regardless of whether it would ultimately change her view. Second, the act of writing an opinion compels each judge to pass on the content of her observation (and interpretations thereof) to successors, even if (once again) these observations are insufficient to overcome the weight of preceding decisions. Such serial communication (if done truthfully¹⁰¹) ensures that each judge contributes to the public inventory of knowledge for future judges, increasing the likelihood that the system can break out of an imprudent cascade at some later date.

In a similar vein, a cascade theory of precedent may shed light on the constraints that many jurisdictions place on *when* a prior opinion can constitute authority. Most federal jurisdictions, for instance, do not accord unpublished (or de-published) opinions any precedential weight whatsoever, and forbid litigants from citing such opinions in written or oral argument.¹⁰² Such a rule comports well with one's anti-herding intuitions: As noted in Part II, the dangers of cascades are greatest when succeeding actors observe only a noisy signal of their predecessors' actions and information (such as the outcome alone). Consequently, the dangers of undesirable cascades are minimized when *either* (i) successors have nearly-complete access to such information, or (ii) they have no information

100. As noted at note 87 *supra*, a cascade can still emerge even when the signals of predecessors are observable, so long as the decision about whether to observe the signal is endogenous.

101. In the example analyzed in Part II, truthful revelation in a written opinion is perfectly consistent with equilibrium play. Outside of information cascade environments, however (such as reputational conformism, discussed in Part IV.C *infra*), truthful opinions may be less plausible.

102. Among federal circuits, those prohibiting citation to unpublished opinions include the First, Second, Seventh, Ninth, and D.C. Circuits. See 1ST CIR. R. 36.2(b)(6); 2D CIR. R. 0.23; 7TH CIR. R. 53(b)(2)(iv); 9TH CIR. R. 36-3; D.C. CIR. R. 28(c). Similarly, the Third Circuit while not formally prohibiting the citation of unpublished opinions apparently adheres to an historical norm against such practices. See 3D CIR. INTERNAL OPERATING PROC. 5.8. (stating that the Circuit "historically has not regarded unpublished opinions as precedents that bind the court," and therefore "the court by tradition does not cite to its unpublished opinions as authority").

For greater elaboration of the federal circuit-court rules on nonpublication and noncitation, see Martha J. Dragich, *Will the Federal Courts of Appeals Perish if They Publish? Or Does the Declining Use of Opinions to Explain and Justify Judicial Decisions Pose a Greater Threat?*, 44 AM. U. L. REV. 757, 761-62 & nn.12-17 (1995); Gregory C. Sisk, *The Balkanization of Appellate Justice: The Proliferation of Local Rules in the Federal Courts*, 68 U. COLO. L. REV. 1, 13-15 (1997).

whatsoever. As a general matter, opinions that remain unpublished are typically much less nuanced and rigorous than are their published counterparts.¹⁰³ Consequently, the relative opacity of unpublished opinions suggests that the herding dangers in using such opinions as authority are potentially much more severe than with their published counterparts. Absent a comprehensive, published opinion, the next best solution for avoiding an undesirable cascade may simply be to disallow an unpublished opinion from having any weight whatsoever, which is exactly the result that nonpublication and noncitation rules (within at least most circuits¹⁰⁴) compel.

In addition to existing judicial practices already mentioned, the information cascades literature adds interesting historical insights on past jurisprudential protocols. For instance, up until the mid-twentieth century, the United States Supreme Court employed an informal voting protocol that had interesting cascade-dampening characteristics. During its regular conferences while in session, the Court votes on (among other things) the disposition of cases argued before it during the previous days. As described by Justice Tom Clark in 1956, these votes traditionally had an interesting counter-seniority flavor: “Ever since John Marshall’s day, the formal vote begins with the junior Justice and moves up through the ranks of seniority, the Chief Justice voting last. Hence the juniors are not influenced by the vote of their elders!”¹⁰⁵

If tenure on the bench proxies (even indirectly) for some measure of “judicial expertise,” then reverse-seniority rules such as that described by Justice Clark would tend to reduce the dangers of herding effects among the judges themselves. Indeed, under such a rule, the late-voting senior Justices—armed with greater relative confidence in their own private

103. William L. Reynolds & William M. Richman, *The Non-Precedential Precedent—Limited Publication and No-Citation Rules in the United States Courts of Appeals*, 78 COLUM. L. REV. 1167, 1175 (1978).

104. It is important to note that the federal circuits, while all disfavoring the citation of unpublished opinions, have not been uniform in their policies towards that end. The Eighth, Tenth, and Eleventh Circuits have a formal policy that such opinions are not binding, but nonetheless authorize citation of such opinions as “persuasive” authority. See 8TH CIR. R. 28A(k); 10TH CIR. R. 36.3; 11TH CIR. R. 36-2. The Fourth and Sixth Circuits are similar, announcing that such citations are officially “disfavored” but nonetheless permissible when the unpublished opinion “has precedential value in relation to a material issue in a case and that there is no published opinion that would serve as well.” See 4TH CIR. R. 36(c); 6TH CIR. R. 10(f). Finally, the Fifth Circuit, while moving towards a policy of disapprobation, still considers any unpublished opinions issued before 1996 to have precedential authority (while those after 1996 carry persuasive authority). See 5TH CIR. R. 47.5.3-47.5.4.

105. Justice Tom Clark, *The Supreme Court Conference*, 19 F.R.D. 303, 306-07 (1956) (emphasis in original).

inclinations—might still follow their own inclinations rather than those expressed by predecessors in action. Conversely, a conventional seniority voting rule poses somewhat greater dangers of queue-taking by junior Justices—who not only may be less confident in their independent assessments of the matter, but also have relatively large reputation-building concerns upon joining the court. Ironically, in the years since Justice Clark penned his account of Supreme Court practice, the reverse-seniority protocol has apparently fallen from use.¹⁰⁶ But whatever the reasons for this development,¹⁰⁷ cascade theory embodies what was almost certainly at the core of the original reverse-seniority voting norm (at least as Justice Clark understood it).

106. Indeed, it appears the practice had already evolved into little more than institutional lore by the early 1970s, as Justice Rehnquist writes:

For many years there has circulated a tale . . . that the voting begins with the junior justice and proceeds back to the Chief Justice in order of seniority. I can testify that, at least during [my tenure] on the Court, this tale is very much of a myth; I don't believe I have ever seen it happen at any of the conferences I have attended.

WILLIAM REHNQUIST, *THE SUPREME COURT* 289-90 (1987).

107. A number of reasons are possible. One obvious candidate is that cascades are simply not a significant danger when the vote is preceded by a period of discussion (as the case with such conferences). It should be noted, however, that under current practices, the vote and the discussion are essentially bundled together. *See* ROBERT L. STERN, EUGENE GRESSMAN & STEPHEN M. SHAPIRO, *SUPREME COURT PRACTICE* 230 (7th ed. 1993) (noting that under current practices, each Justice customarily indicates his or her vote at the first opportunity to comment). Consequently, the only formal discussion that the Chief Justice hears before casting his vote is essentially his own. A related but slightly different theory is that a reverse-seniority voting rule is not likely to stem cascading phenomena that have already begun in pre-vote deliberations. Indeed, even during Justice Clark's tenure on the bench (as today), *discussion* of argued cases within conference sessions occurred sequentially, in order of seniority starting with Chief Justice and ending with the most junior Associate Justice. *See id.* at 6-7. Given that late contributors may have already distorted their expressed viewpoints based on what more senior justices have said, the reverse-seniority protocol at a subsequent voting stage may not present much of a safeguard against cascades. Yet a third possible reason that reverse-voting protocols have not survived is the observation that increased workload on the Court has caused it to streamline its procedures to save time, eliminating a separate stage of voting at the end. *See id.* at 29-38. Finally, senior Justices may have strong personal incentives to foster a conventional seniority norm, as reverse-seniority rules wrest some agenda-control powers from the top of the Court's hierarchy. As a Justice moves up through the ranks of seniority, she or he may increasingly feel uninterested in advocating an institutional protocol that disfavors senior members of the Court. *See, e.g.,* REHNQUIST, *supra* note 106, at 289 (describing his disappointment while an Associate Justice that his judicial efforts were rarely heeded, as the outcome of a vote was already determined by the time his opportunity to speak and vote arrived, and noting that as his tenure increased on the Court, he came to realize that his preference for more of a round-table discussion was fine in the abstract, "but was ultimately Adomed by the seniority system to which the senior justices naturally adhere"). *Accord* Ruth Bader Ginsburg, *Essay: Remarks for American Law Institute Annual Dinner*, 38 ST. LOUIS L.J. 881 (1994) (relating the same story, but noting that one redeeming virtue of being the last voter is the possibility of being the deciding vote in a 5-4 decision).

B. CASCADES AND NORMATIVE ANALYSIS

Beyond enriching our positive understanding of existing (and/or historical) institutions, appreciating the possibility for precedential cascades may also provide some limited normative guidance to judges in at least two ways. First, even though existing organizational devices in the judiciary tend to dampen the likelihood of cascades *among judges*, similar institutional devices need not constrain the behavior of nonlegal actors. As such, cascade theory can still be of enormous value in helping judges to understand the populations they endeavor to regulate. It has been suggested, for instance, that public assessments of catastrophic environmental risks may be subject to cascade phenomena, which, if left unchecked, can lead to equally catastrophic policy decisions.¹⁰⁸ Financial markets may similarly be subject to herding behavior, leading to market bubbles and concomitant panics or crashes.¹⁰⁹ When a judge is attempting to articulate legal rules to regulate such actors, she would be wise to assess whether and to what extent herding dangers exist. When they are likely, judges may be able to stem such events, by (inter alia) mandating the public disclosure of information that might break a cascade, constraining the behavior of those who might generate one,¹¹⁰ or by effective oversight of administrative agencies (who may themselves fall victim to such phenomena).¹¹¹

Second, cascade theory can potentially provide a normative compass (albeit a limited one) in assessing the prudence of other legal institutions and/or proposals. In their influential book on the structure of judiciary, for example, Sam Estreicher and John Sexton advocate a reform of the Supreme Court's case selection criteria using a quasi-managerial model of docketing.¹¹² Under this approach, the Supreme Court should privilege the

108. See Kuran & Sunstein, *supra* note 10.

109. See, e.g., David Hirshleifer, *Informational Cascades and Social Conventions*, in THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 301, 305-06 (citing sources).

110. For instance, Section 16(a) of the Securities Exchange Act of 1934 requires corporate fiduciaries and large-stake (10% or higher) shareholders to identify themselves to the SEC and report any significant trades they make in the market for that security. Moreover, Section 16(b) requires such insiders to disgorge to the issuer any "short-swing" profits to the issuer made within 6 months of a purchase and sale of the proscribed security, thereby dampening their incentives to ignite a cascade (or alternatively inducing them to disclose additional information to the market).

111. Kuran and Sunstein, for example, suggest that courts may be able to use their powers under the Administrative Procedure Act to reverse "arbitrary and capricious" administrative regulations which may be the product of a public risk-assessment cascade. See Kuran & Sunstein, *supra* note 10, at 758-59.

112. SAMUEL EISTREICHER & JOHN SEXTON, REDEFINING THE SUPREME COURT'S ROLE 41-70 (1986).

existence of persistent, inter-court conflict when making certiorari decisions.¹¹³ By negative corollary, of course, this proposal impels the Court to place relatively little weight on areas of law where there is complete or substantial harmony among competing circuits or state supreme courts.¹¹⁴

Significantly, the Estreicher-Sexton justification for privileging inter-jurisdictional conflicts is not merely academic: The Supreme Court's own rules on certiorari make it clear that inter-jurisdictional conflict is a principal guiding factor in reaching a decision:

A review on writ of certiorari is not a matter of right, but of judicial discretion. A petition for a writ of certiorari will be granted only when there are special and important reasons therefor. The following, while neither controlling nor fully measuring the Court's discretion, indicate the character of reasons that will be considered:

- (a) When a United States court of appeals has rendered a decision in conflict with a decision of another United States court of appeals on the same matter; or has decided a federal question in a way in conflict with a state court of last resort
- (b) When a state court of last resort has decided a federal question in a way that conflicts with the decision of another state court of last resort or a United States court of appeals.¹¹⁵

The "if-it-ain't-broke-don't-fix-it" rationale that lies at the core of the inter-jurisdictional conflict criterion certainly seems unobjectionable at first glance. Indeed, the Supreme Court has limited resources, and can only hear a fraction of appealed cases each year.¹¹⁶ Thus, the Court would certainly be well-advised to focus only on only those strands of case law in which its intervention is seriously needed.¹¹⁷ The pivotal question remains, however, whether inter-jurisdictional conflict provides a reliable diagnostic

113. To animate their reform proposal, Estreicher and Sexton analogize the job of the Supreme Court to that of a manager who oversees a hierarchy of delegated decisionmakers:

A wise manager delegates responsibilities to subordinates and, when there is no indication that something is awry, does not intervene. To do otherwise is to denigrate the authority of subordinate actors, diminishing their own sense of responsibility and ultimately increasing the manager's tasks as well as the overall workload. The Supreme Court as manager would trust the subordinate actors in the judicial system, intervening *only when some structural signal (such as a persistent conflict between subordinates) indicates a problem requiring correction.*

Id. at 50 (emphasis added).

114. Indeed, Estreicher and Sexton label as "improvident" the granting of certiorari to areas of law in where only two or fewer lower courts are in doctrinal conflict. *See id.*

115. SUP. CT. R. 10.

116. *See STERN ET AL.*, *supra* note 107, at 29-38.

117. A similar argument may well hold true for appellate-level courts, which, while necessarily taking on all appeals, may decide on the degree of intensity with which they review.

proxy for such situations. A serious account of precedential cascades raises the possibility that such a criterion is severely under-inclusive. In particular, if judges within sister jurisdictions viewed one another's opinions as potentially persuasive authority, cross-jurisdictional uniformity *might* mask an underlying cascade—and with it, stagnated judicial learning and possibly serious errors in policy. In contrast, areas of case law in which there is significant inter-jurisdictional division might actually portend *good news* (at least in some cases), as it indicates that dynamic judicial learning may still be vibrant and ongoing. Hence, rational-herding considerations may conflate considerably one's normative assessment of the intra-jurisdictional conflict criterion.¹¹⁸

It is, however, important to remain mindful of a limiting caveat to such normative arguments: the aforementioned difficulty in empirically verifying cascades¹¹⁹ would almost surely confound the Court's attempts to distinguish (a) the uniformity produced by cascades, and (b) that produced by more generic intra-jurisdictional harmony. Although the Court might attempt to marshal some set of prescriptive considerations (for example, the degree to which various opinions cite one another, the relative opacity in the opinions, or the degree to which prudent policy turns on information learned through the cases¹²⁰), it could never be completely sure that its intervention in a particular strand of cases has positive net value. But by the same token, the very difficulty in proving the *existence* of cascades also implicates the task of proving their *absence*. As such, the potential danger of disguised uniformity should perhaps play at least some role in case selection, and even an occasional, seemingly improvident review may be justifiable given the possible costs of imprudent precedential herding.

C. NON-LEARNING-BASED SOURCES OF CONFORMISM

Finally, the arguments presented above may help to constrain and focus the broader debate about whether precedent represents a dangerous form of group herding. While information cascades may not be a plausible source of judicial conformity, they are but one positive account of

118. To be sure, there may be other reasons (not related to cascades) for courts to close up inter-jurisdictional disagreements (at least eventually). At some point the value of uniformity and predictability in law likely surpasses the value of additional fine-tuning. Viewed in this sense, it may well be prudent for the Court to step in and resolve inter-jurisdictional conflict after sufficient percolation. Nevertheless, the point remains that if the Court is also interested in attenuating precedential cascades, such a criterion is almost surely under-inclusive.

119. See Part III, *supra* notes 79-98.

120. See, e.g., Daughety & Reinganum, *supra* note 10.

conformity. Although a comprehensive analysis of alternative accounts is beyond the scope of this Article, it is perhaps worthwhile to take up briefly two such possibilities: (a) reputational concerns, and (b) the role of cognitive biases towards conformity.¹²¹ Both of these phenomena deserve consideration as possible independent sources of dangerous herding behavior. But at the same time, a number of open questions remain as to the independent power and plausibility of these alternative hypotheses on their own terms.

First, individual judges may refuse to break rank with their predecessors because of a *reputational stake in conforming*, perhaps for fear of subjecting their own preferences or intellectual shortcomings to public scrutiny should they break with the herd. A judge may, for example, have a large stake in bolstering her standing within some public constituency, and might therefore be inclined to issue holdings that favor that group regardless of her own inclinations.¹²² Alternatively, a judge may fear that he is not as skilled as his predecessors, and may therefore attempt to “hide” within existing precedent, believing that if the policy it represents is eventually exposed as improvident, it is much better to fail within a group than to do so alone.¹²³ Over time, these reputational concerns may give rise to a stable precedent, but one that is not justified on the basis of sequential learning.

Reputational stories such as this are potentially quite powerful, and therefore warrant attention from those interested in the formation and durability of judicial conformity. Such accounts, however, inevitably must confront many of the same shortcomings that rendered standard information cascades somewhat implausible. For instance, the degree to which a judge is beholden to a particular constituency seems much diminished when that judge enjoys life tenure (as do most federal Article III judges and some state supreme court judges). In fact, as one moves through the appeals process (and in the limit to the U.S. Supreme Court),

121. Omitted from the textual discussion, but nonetheless relevant, are so-called network effects, in which individuals' payoffs from adhering to a given practice increases with the population of others who adhere to the same practice. Of particular relevance here are computer programs and hardware, telephones, and trade practices. *See supra* cites at note 13. The role of network effects is helpful in understanding why some conformity trends may eventually become stable, but it tells us little about how or why the initial convention starts, and why at its nascent stages players would want to conform with one another. The alternative theories presented in the previous section and below attempt to offer such a theory, while network effects do not.

122. *See* Kuran & Sunstein, *supra* note 10, at 727-30. *See also generally* TIMUR KURAN, PRIVATE TRUTHS, PUBLIC LIES: THE SOCIAL CONSEQUENCES OF PREFERENCE FALSIFICATION (1995).

123. *See* Zwiebel, *supra* note 12; Scharfstein & Stein, *supra* note 12.

the constituency-serving proclivities of those on the bench likely recede even further. Thus, even if reputational herding occurs within the lower echelons of the judiciary, the appeals process itself may be a particularly good institutional response to such dangers.

Moreover, the sorts of equilibria that come from reputational attempts to hide quality differences are subtle, and need not always entail group conformity. In one well-known application of reputational herding to corporate managers, for example, Jeff Zwiebel argues that extremely high-quality managers are likely to *break* with convention and opt for riskier but more profitable projects.¹²⁴ In fact, his analysis predicts that an extremely *low*-quality manager will also tend to break from convention, hoping against hope that a successful outcome may lead outside observers to reassess his quality, thereby prolonging his career. In the context of legal precedent, a similar outcome might occur, in which very high or very low quality judges are willing to break with current law *precisely because of* reputational concerns. Thus, just as with information cascade models, heterogeneity among the actors in a reputational model can quite plausibly reduce (rather than enhance) a trend towards conformity.

Another strand of behavioral hypotheses that can lead to precedent-like conformity comes from the field of cognitive psychology. For many years, experimental psychologists have made significant contributions to social science by demonstrating that individuals often deviate in important ways from rational behavior in environments of risk or uncertainty. Sometimes these deviations are severe, and undoubtedly implicate important policy concerns when they occur outside of the laboratory. Indeed, Timur Kuran and Cass Sunstein have recently posited that one such behavioral pathology—known as the “availability heuristic”—may lend itself to cascade-like behavior.¹²⁵

The availability heuristic is a cognitive bias in the way people infer the probability of some event from their own experiential observations about how frequently it occurs. At its core, the heuristic asserts that one’s empirical recollections (which comprise her statistical “sample”) will be skewed towards particularly memorable or vivid examples, giving relatively short Bayesian shrift to more mundane—but perhaps more representative—occurrences.¹²⁶ As applied to precedent, the availability

124. See Zwiebel, *supra* note 12.

125. See Kuran & Sunstein, *supra* note 10.

126. See Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124 (1974); Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 4 COGNITIVE PSYCHOL. 207 (1973).

bias might induce judges to place excessive decisional weight on a few well-remembered holdings (such as those involving particularly recent disputes, or prominent litigants or judges) when forming their estimations of sound legal policy. By placing such great emphasis on a relatively modest sample, the judge may well decide to follow the precedent from those cases even if the majority of less vivid cases came out the other way.

While certainly pertinent and worthy of investigation, the availability heuristic faces at least three challenges if it is to carry the weight of a cascade theory of precedent. First, it may just as plausibly work *against* a norm of precedent than for it. For instance, in deciding whether to follow precedent and impose a mandatory life prison term on a convicted criminal defendant, a judge may systematically recall only the most vivid instances in which such defendants were much later shown to be innocent (even if most sentences are handed down appropriately).¹²⁷ Alternatively, she may selectively recall only those cases that *she personally* adjudicates, since those are the only ones whose facts are particularly vivid for her. Once again, such behavior could tend to dampen—not encourage—mutual emulation among judges.

Second, the availability heuristic is but one of a vast array of nonrational cognitive biases identified within prospect theory. Other biases may be equally plausible, yet work at cross purposes with the availability heuristic. Consider, for example, the so-called “representativeness” heuristic, which posits that when analyzing a population sample to choose between two hypotheses about that population, decision makers tend to ignore a priori information they inherit about base-rate probabilities of the two hypotheses. In the context of the example from Part II, this implies that a judge’s choice between the two relevant hypotheses (that is, whether most of the population consists of low-cost or high-cost victims) would turn predominantly on the observed cost of the victim in the case before her, *placing little weight on the information that prior holdings might convey about the population in general*. Once again, such a tendency appears to mitigate (rather than exacerbate) the dangers of undesirable cascades.

Similarly, the oft-studied cognitive phenomenon of overconfidence may also have cascade-dampening effects. Psychologists have demonstrated that a vast majority of people rate themselves as more

127. An identical argument would apply, of course, if the received precedent were relatively lenient, as the judge might vividly recall the few instances of a culpable defendant who was released early and proceeded to commit more crimes.

intellectually gifted than others, better drivers than the average person,¹²⁸ and less likely than the average person to fall victim to a fatal disease.¹²⁹ Newlyweds generally estimate correctly that approximately half of all married couples eventually divorce, but at the same time assess their own chances for divorce to be zero.¹³⁰ Such “Lake Wobegon” effects appear to implicate the judgment of professionals and entrepreneurs as well, and transcend a number of other demographic profiles.¹³¹ It would seem to be particularly salient among judges as well, who are often (but not always) selected for the job on the basis of their demonstrated skills as high-profile lawyers or professors.¹³² The possible dampening effect of overconfidence on the formation of precedential cascades is therefore apparent: Overconfidence induces the marginal actor to *discount* (rather than inflate) the relative decisionmaking skills of her counterparts, thereby inducing her to be less inclined to follow the precedents established by those perceived to be less talented actors than herself.¹³³

Finally, the possibility that precedent arises through non-information-based cascades does not ease the burden of verifying whether such conformity is *caused* by such possibilities, or rather by other factors that simply are not observed by the statistician. In fact, cognitive accounts may be similarly prone to verifiability problems outside the laboratory, as most of our current stock of knowledge consists more of a set of controlled

128. See, respectively, 2 R.C. WYLIE, *THE SELF-CONCEPT* (1979); O. Svenson, *Are We All Less Risky and More Skillful Than Our Fellow Drivers?*, 47 *ACTA PSYCHOLOGICA* 143 (1981).

129. See N. Weinstein, *Unrealistic Optimism About Susceptibility to Health Problems*, 5 *J. BEHAV. MED.* 441 (1982). This sort of over-optimism would also appear to implicate how people react to public health crises, and might therefore tend to counteract the sort of public scare phenomena described in Kuran & Sunstein, *supra* note 10.

130. See Lynn A. Baker & Robert E. Emery, *When Every Relationship Is Above Average: Perceptions and Expectations of Divorce at the Time of Marriage*, 17 *LAW & HUM. BEHAV.* 439, 443 (1993).

131. Colin F. Camerer & Howard Kunreuther, *Decision Processes for Low Probability Events: Policy Implications*, 8 *J. POL'Y ANALYSIS & MGMT.* 565, 569 (1989); Neil D. Weinstein, *Optimistic Biases About Personal Risks*, 246 *SCI.* 1232, 1232 (1989). The label “Lake Wobegon effect” refers to Garrison Keillor’s mythical city on the Prairie Home Companion radio show “a city where “the women are strong, the men are good looking, and all the children are above average.”

132. It has often been noted that the legal profession is one of the most salient examples of so-called “winner-take-all” markets, which tend to be inhabited by habitually overconfident individuals. See, e.g., ROBERT FRANK & PHILIP COOK, *THE WINNER-TAKE-ALL SOCIETY* 16-17, 97-98 (documenting this feature within the legal profession).

133. 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).

experimental results than it does a unified, robust theory of behavior.¹³⁴ And while this does not mean that such accounts are “wrong,” it does suggest that more work lay ahead.

As the discussion above demonstrates, reputational and cognitive phenomena potentially play a role in animating a more textured and general account of precedent as a form of herding behavior. If, however, information cascades can play no more than a modest role in such an account (as I have argued in previous sections), these alternative hypotheses must carry greater proportional weight, and must do so amid their own plausibility challenges. Whether they can support the weight of the argument remains to be seen, but is nonetheless worthy of future exploration.

CONCLUSION

Cascades and herding behavior have become increasingly popular tropes among legal academics in recent years.¹³⁵ This trend is hardly surprising: Traditional rational actor models are almost certainly an incomplete account of individual behavior in practice, and have had only moderate success in predicting group behavior. Models of cascade behavior offer an attractive explanation of how groups of individuals may be prone to severe lapses of judgment—lapses which may well be magnified later by other well-documented behavioral pathologies. Whether such phenomena help explain formation and stability of judicial precedent is an interesting and important question. Nevertheless, as the arguments presented above suggest, it seems unlikely that information cascades present a significant impediment to the judiciary. At present, then, the cascades literature does not provide general and robust description 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, a cascade theory of precedent lends considerable justification to a number of existing organizational features of the judiciary as cascade-preventing devices, such as life tenure, written opinions, stacked hierarchies, and the frequent preference for judicial casuistry over hard-and-fast rules.

134. See Jennifer Arlen, *The Future of Behavioral Economic Analysis of Law*, 51 VAND. L. REV. 1765, 1787-88 (1998) (noting that in spite of its attraction, prospect theory has yet to offer a coherent, robust alternative paradigm to the admittedly-flawed rational choice paradigm).

135. A recent LEXIS search found well over 100 published law review articles that refer to either “herding behavior” or “information cascades” or both since 1992.

Moreover, cascade theory can still be of enormous value in helping judges to understand the populations they endeavor to regulate—populations that are not themselves governed by the same cascade-dampening institutions that protect the judiciary. Additionally, awareness of the possibility of cascades may provide at least a small amount of normative guidance in evaluating the optimal design of other institutions, such as the Supreme Court’s method of selecting cases for review. And finally, the implausibility of precedential cascades nonetheless helps focus and refine future debate around alternative non-informational theories of dangerous judicial conformity. Although these alternative theories face plausibility and verifiability challenges of their own, the enterprise of exploring them clearly warrants our cautious attention.¹³⁶ Legal scholars, judges, and practitioners would be wise to heed the lessons from the cascades literature in considering the role of precedent.¹³⁷ But in so doing, it is important to be sure that it is not we—as opposed to those we study—who are the ones caught within a cascade.

APPENDIX

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	=	Judge in period i
c	=	Injurer’s cost of avoidance (constant for all injurers)
x_i	=	Signal received by J_i about victim (1 if high-cost, 0 if else)
$c + d(x)$	=	Randomly-drawn victim’s cost of avoidance ($d(1) = D > 0$ if victim is high-cost, and $d(0) = -D < 0$ if victim is low-cost)

136. See, e.g., Kuran & Sunstein, *supra* note 10. 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., Bernardo & Welch, *supra* note 133.

137. Contract interpretation doctrine, for instance, often hinges on established trade custom. See, e.g., U.C.C. § 2-202.

Y_i	=	J_i 's announced holding (1 if pro-victim, 0 if pro-injurer)
θ	=	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:

$$c < \int_0^1 [(c + D) \cdot \theta + (c - D) \cdot (1 - \theta)] dF(\theta / x_i, h_i)$$

Which, after simplification, is equivalent to:

$$\int_0^1 (\theta) dF(\theta / x_i, h_i) < \int_0^1 (\theta) dF((1 - \theta) / x_i, h_i)$$

The above expression will always be satisfied under first-order stochastic dominance; that is, $F(\theta / X_i, h_i) < F(1 - \theta / X_i, h_i)$ for all $\theta \in [0, 1]$.

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

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

Proof: "IF": Under myopia, where $h_i = \{\emptyset\}$, 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 θ (recalling the uniform prior) will then be:

$$\begin{aligned} f(\theta / X_1 = 1) &= \frac{\Pr\{X_1 = 1 / \theta\} \cdot f(\theta)}{\int_0^1 \Pr\{X_1 = 1 / \theta\} \cdot d\theta} = \frac{\theta}{\int_0^1 \theta d\theta} \\ &= 2\theta \end{aligned}$$

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

Proposition 2: Suppose instead that judge J_i has access to both previous holdings and the factual record that drove those opinions (that is, $h_i = \{y_1, y_2, y_3, \dots, y_{i-1}, x_1, x_2, x_3, \dots, x_{i-1}\}$). Judge J_i will prefer $y_i = 1$ ($y_i = 0$) whenever the total number of $x = 1$ signals exceeds than (is less than) the number of $x = 0$ signals.

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 θ given n_0 and n_1 is given by Bayes rule as follows:

$$\begin{aligned} f(\theta | n_0, n_1) &= \frac{\Pr\{n_0, n_1 | \theta\} \cdot f(\theta)}{\int_0^1 \Pr\{n_0, n_1 | \theta\} \cdot d\theta} = \frac{(1-\theta)^{n_0} \cdot \theta^{n_1}}{\int_0^1 (1-\theta)^{n_0} \cdot \theta^{n_1} d\theta} \\ &= \frac{(n_0 + n_1 + 1)!}{n_0! n_1!} \cdot (1-\theta)^{n_0} \cdot \theta^{n_1} \end{aligned}$$

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

$$\left[\frac{\theta}{1-\theta}\right]^{n_1} > \left[\frac{\theta}{1-\theta}\right]^{n_0}$$

This implies that when $n_1 > n_0$, $f(\theta | n_0, n_1)$ crosses $f(1-\theta | n_0, n_1)$ from below but one time, and satisfies the single crossing property for probability density functions.¹³⁸ This ordering is reversed whenever $n_1 > n_0$. As such, $F(\theta | n_0, n_1)$ dominates $F(1-\theta | n_0, n_1)$ 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.

138. See Paul R. Milgrom, *Good News and Bad News: Representation Theorems and Applications*, 12 BELL J. ECON. 380 (1981).

Proposition 3: Suppose historical holdings are opaque (i.e., $h_i = \{y_1, y_2, y_3, \dots, y_{i-1}\}$). 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 $\{y_1, y_2, y_3, \dots, y_{i-2}\}$ perfectly reveals the content of the signals $\{x_1, x_2, x_3, \dots, x_{i-2}\}$: half $(i-2)/2$ courts received signal $x=0$ and half $(i-2)/2$ received signal $x=1$.

Without loss of generality, suppose that $y_{i-1} = y_i = 1$.¹³⁹ 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 $y_{i+1} = 1$ even when $x_{i+1} = 0$.¹⁴⁰ Thus, suppose $x_{i+1} = 0$. J_{i+1} 's posterior density on θ is given by:

$$\begin{aligned} f(\theta \mid h_i, x_{i+1}=0) &= \frac{[\theta + (1-\theta)\alpha] \cdot (1-\theta)^{i/2} \theta^{i/2}}{\int_0^1 [\theta + (1-\theta)\alpha] \cdot (1-\theta)^{i/2} \theta^{i/2} d\theta} \\ &= K_i \cdot [\theta + (1-\theta)\alpha] \cdot (1-\theta)^{i/2} \theta^{i/2}, \end{aligned}$$

where K_i is a constant.¹⁴¹ Using the expression above, it is easily verified that $f(\theta/h_i, x_i=0) < f(1-\theta/h_i, x_i=0)$ if and only if $\theta < 1/2$, thereby satisfying the single crossing property.¹⁴² As such, θ dominates $(1-\theta)$ in the sense of

139. The case of $y_{i-1} = y_i = 0$ is identical, and is therefore ignored.

140. Certainly, if one can show this result, then it is straightforward to show the weaker result that $y_{i+1} = 1$ if $x_{i+1} = 1$.

141. Explicitly, K_i is given by:

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

142. See Milgrom, *supra* note 138.

first order stochastic dominance. Importantly, once this condition is verified for J_i , 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 $n_i \geq n_0+2$ 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.