Judicial Fact Discretion[1]

Nicola Gennaioli and Andrei Shleifer

Second Draft, July 2006

Abstract

Judicial fact discretion is defined as misrepresentation in a judge's decision of facts revealed in a trial. We identify two motivations for the exercise of such discretion: judicial policy preferences and judges' aversion to reversal on appeal when the law is unsettled. In a standard model of a tort, we show that judicial fact discretion leads to distorted incentives for taking precautions, excessive levels of bad accidents, trial outcomes unpredictable from true facts of the case, and welfare losses. It also raises the incidence of litigation relative to settlement, and encourages litigants to take extreme positions in court. The model makes plausible predictions as to when common law trials will work poorly, and points to legal rules limiting distortions from judicial fact discretion.

1. Introduction

In his "Courts on Trial," the distinguished American legal realist Jerome Frank defined the concept of judicial fact discretion:

When the oral testimony is in conflict as to a pivotal fact-issue, the trial judge is at liberty to choose to believe one witness rather than another. In other words, in most cases the trial judges have an amazingly wide "discretion" in finding the facts, a discretion with which upper courts, on appeals, seldom interfere, so that, in most instances, this "fact discretion" is almost boundless. … As one court put it , "the word 'discretion' is properly enough used to express the judicial judgment in discriminating as to weight and cogency between different witnesses… which must be exercised in reaching any conclusion of fact from evidence." (Frank 1951, p. 57).

Frank recognized that some fact discretion is unavoidable, since judges necessarily have to decide which witness accounts to trust. But Frank (and later Posner 2005) also recognized that fact discretion creates significant leeways for the expression of judicial preferences, which derive from political, social, or economic views, or even from a judge's career concerns. To Frank, fact discretion is a major source of unpredictability of judicial decisions, since it allows judges to express their sentiments through fact finding rather than the application of the law.

In this paper, we introduce judicial fact discretion into a formal analysis of trial court decision making[2]. To this end, we identify two distinct motives for the exercise of judicial fact discretion. The first, emphasized by legal realists, is judicial bias. As Posner (2005, p. 14) – echoing Frank (1930, 1932) – writes about federal district (i.e., trial) judges: "But [deciding a particular case in a particular way might increase the judge's utility] by advancing a political or ideological goal, economizing on the judge's time and effort, inviting commendation from people whom the judge admires, benefiting the local community, getting the judge's name in the newspaper, please a spouse or other family member or a friend, galling a lawyer whom the judge dislikes, expressing affection for or hostility toward one of the parties – the list goes on and on."[3]

---

[2] Although our model focuses on judges, much of the discussion – particularly the sections focused on judicial bias -- is applicable to juries as well. Strictly speaking, our model deals with fact-finder fact discretion.

[3] Gennaioli (2004) formally introduces judicial bias into a model of adjudication. Gennaioli and Shleifer (2006) and Ponzetto and Fernandez (2006) examine the consequences of appellate courts' bias for the evolution of the law. Mullainathan and Shleifer (2005) model a similar kind of "fact discretion" by the media in its reporting of the news.

The second motive, specific to trial judges, is the dislike of being overruled by appellate courts. As Posner (2005, p.16) comments: "Judges also don't like to be reversed, even though a reversal has no tangible effect on a judge's career if he is unlikely to be promoted to the court of appeals in any event." Appellate courts typically do not revisit facts found by trial courts, but only the application of the existing law to those facts. When such application is uncertain, a trial court has an incentive to "fit" the facts into the settled precedent, so that from the point of view of the appellate court, the application of the law to the facts is uncontroversial.

We consider each of these motives for the exercise of judicial fact discretion in a standard model of a tort. In this model, the first best efficient legal rule is strict liability with all harms being legally cognizable for the calculation of damages. We assume that trial courts follow this rule, but can distort facts about harm. We show that fact discretion renders judicial decisions less responsive to the facts of the case than is socially optimal in both the judicial bias (section 3) and the fear of overruling (section 4) models. As a consequence, judicial fact discretion distorts the taking of precautions, leading in particular to a higher incidence of very harmful accidents. It renders judicial decisions unpredictable, and reduces welfare. In section 5, we further show that judicial fact discretion increases the amount of litigation as opposed to settlement, and encourages litigants to take extreme positions in court, especially in the instances where the law is unsettled and judges are motivated to distort fact finding to avoid reversal. Taken together, these results suggest that the legal system is likely to perform poorly in the areas of law that are fact-intensive, relatively new (so precedents are undeveloped), and political in the sense that judicial preferences play a large role. The highly controversial product accident litigation may fit this description.

In section 6, we turn this problem around and ask which legal rules are optimal in the presence of judicial fact discretion. We show that strict liability with all harms being legally cognizable for the calculation of damages remains the optimal rule when relatively few judges are biased. On the other hand, when enough judges are biased, it may be efficient to restrict the range

of legally cognizable harms, and therefore the flexibility for judges to set damages, so as to avoid the adverse consequences of the exercise of fact discretion. This analysis might shed light on some restrictions the law places on cognizable harms, such as the economic loss doctrine.

Legal scholarship is a bit schizophrenic on the subjects of judicial fact finding and discretion. On the one hand, practicing attorneys (including those teaching at law schools), as well as popular writers on the law, see the sympathies, the mood, and the psychology of a judge as central to determining the outcome of a trial. A distinguished trial judge writes that "there is a terrific importance in the trial court, never equaled in any appellate court, of knowing who is the judge" (Wyzanski 1972, p. 208). On the other hand, traditional legal scholarship has ignored Frank's thinking. According to Brunetti (1998, p. 1493) "much has been written about judicial philosophy and almost nothing has been written about the fact-finding aspect of trial court decision making… An entire bibliography published in 1993 of every book and article ever written about judicial decision-making contains no source on fact-finding by trial courts." Some recent work in law and economics examines the consequences of trial court error (e.g., Kaplow 1994, Kaplow and Shavell 1994, 1996), but even this work does not deal with the central issue of judicial motivation underlying error.

An Example of Judicial Fact Discretion

Before turning to the formal analysis, we present an example of the exercise of judicial fact discretion in one famous case. We note, first, that fact discretion has been studied in the context of appellate courts, where the idea has been that appellate judges sometimes "simplify" the facts to elucidate a legal principle. A very clean example of this is Cardozo's extreme mischaracterization of facts in *MacPherson v. Buick* (Henderson 2003), but Cardozo appears to have altered the facts at least marginally in *Palsgraf* as well (Posner 1990). Dershowitz and Ely

(1971) denounce the Burger Supreme Court for its extreme mischaracterization of facts in *Harris v New York*, an exercise of fact discretion that the authors call "the failure of candor."

Our example of judicial fact discretion in action comes from one of the first cases in the standard torts textbook (Keeton, Sargentich, and Keating 2004), *Garratt v. Dailey*. In textbooks, the case stands for the proposition that knowledge of possible harm is sufficient to find intent in battery, so the plaintiff does not need to show purpose to harm to establish the defendant's liability. But the case is also a clear instance of judicial fact discretion. Although it does not deal with the exact situation we study, *Garratt v. Dailey* shows a trial judge completely changing his *fact finding* after an appellate court remands the case back to him on a matter of law.

Brian Dailey, a five year old boy, accompanied his mother on a visit to his aunt, Ruth Garratt, in the garden of Garratt's house. The boy allegedly pulled a chair from under his aunt as she started to sit down, she fell and injured herself, and subsequently sued Brian. According to the appellate court review of the evidence, "the trial court accepted the boy's statement that he had moved chair and seated himself therein, but, when he discovered that plaintiff was about to sit at place where chair had been, *attempted to move chair toward plaintiff, and was unable to get it under plaintiff in time*." (cite, italics added) Having accepted the boy's view that he was trying to help his aunt rather than hurt her, the trial court ruled for the boy on the grounds that he did not have the purpose – and therefore intent – to harm her.

The appellate court ruled that purpose to harm is not required to prove intent in battery, and that knowledge of possible harm is sufficient, and remanded the case back to the trial judge (in this case, superior court). "Upon remand for clarification on the issue of the defendant's knowledge, the superior court reviewed the evidence, listened to additional arguments and studied briefs of counsel, and entered a finding to the effect that the defendant knew, with substantial certainty, at the time he removed the chair, that the plaintiff would attempt to sit down where the chair had been, since *she was in the act of seating herself when he removed the chair*." (cite,

italics added)   The trial court shifted all the way from the finding that the boy was moving the chair *toward* the aunt as she was sitting down to the finding that he was pulling it *from under* her.

There may be a number of explanations, some innocent, for how the trial court found such entirely different facts after the case was remanded. But there are two simple stories. First, the judge might have been initially annoyed with the aunt for bringing a case against her 5 year old nephew, presumably to collect insurance, and so accepted the boy's somewhat bizarre testimony to reach his initial verdict. He could have, and of course eventually did, accept the other testimony instead. This judicial bias view of fact discretion is analyzed in Section 3.

Alternatively, the trial judge might have feared reversal. When he thought that the standard of intent in battery was purpose, he found the facts under which the boy could not possibly have had the purpose to harm his aunt, namely that he was moving the chair *toward* her. Under the legal rule the judge believed, the factual finding that the boy was pulling the chair from under his aunt would have raised a difficult question of whether he had the purpose to harm her, and exposed the judge to the risk of reversal if the appeals court took a more benign view on an unsettled legal question of a 5 year old's purpose. To the judge's surprise, the appellate court took a radically different view of the meaning of intent in battery. So when the trial judge learned that the standard of intent was merely knowledge of possible harm, and not purpose to harm, he found the facts under which knowledge was pretty much obvious, even to a 5 year old.   Had he stuck to his old finding of facts that he boy was trying to put the chair back, the question of knowledge would have been legally controversial. In both of his decisions, the trial judge found the facts that render the application of the law that he believed to be in place utterly straightforward. We consider this motivation for fact discretion in Section 4.

2. Judicial Fact Discretion

Consider a tort where injurer *I* harms victim *V*. *I* could be a company using explosives and *V* a resident whose property is damaged in an accident with explosives. *V*'s harm from the injury, denoted by h, is uniformly distributed on $[0, D]$, where $D \leq 1$ is the maximum harm *I* can inflict on *V*. We assume that *I* knows h before he engages in the potentially harmful action: he knows what harm would result if explosives are stored or transported in different ways[4].

At a cost $c(p) = (1/2)p^2$, *I* can take precautions $p \in [0,1]$ to avoid the injury. For example, p could represent the company's effort to transport explosives more securely or to store them further away from *V*'s property. With precautions p, the victim suffers harm only with probability *1-p*. Since the level of harm h is known in advance to the injurer, expected social losses from taking precautions *p* are given by:

$$(1-p)h + (1/2)p^2 \tag{1}$$

First best precautions are then equal to $p_{fb}(h) = h$. Aggregate social losses in the first best (L_{fb}) when *I* takes optimal precautions are equal to:

$$L_{fb} = \int_0^D \left[(1-h)h + (1/2)h^2 \right](1/D)dh = D(3-D)/6 \tag{2}$$

We study torts where there is no contract, or alternatively where it is too costly for the parties to specify precautions in a contract. As in the standard model of torts (Posner 1972, Shavell 1987), the injurer's precautions are shaped by the damages set by courts in light of the prevailing legal rule. For simplicity and in line with the explosives example, we study strict liability regimes in which full harm h is legally cognizable. In the absence of judicial fact discretion, this legal rule delivers the first best efficient level of precautions.

We assume in sections 2-4 that the judge observes h, but can make the finding of fact of

---

[4] In our model, if h is totally unknown in advance, the problem becomes trivial. The analysis goes though if the injurer has some but not perfect information about h.

$h' \neq h$, and thus apply damages $d(h') \neq d(h)$. This is our definition of judicial fact discretion: a court's finding of harm *h'* does not need to coincide with the facts *h* revealed in a trial.

There are several ways in which judges may be able to "work on the facts" of a case. Some of the evidence presented to them may be oral, and so they may choose whom to believe. The documents in the evidence may include ambiguous language, which judges are free to interpret. The victim's harm may depend on a multitude of potentially conflicting factors. By emphasizing certain pieces of evidence and neglecting others, a judge may discretionally alter the facts of the case to meet his desired level of harm. In general, the extent of fact-discretion is likely to depend on physical attributes of the transactions (e.g., the number of empirical dimensions material to *h*), which determine the private cost to the judge of distorting the facts. However, to see the implication of fact discretion in the simplest and starkest way, we assume that it is *absolutely costless* for judges to bend the facts of a case. This does not imply that the extent of judicial fact discretion is unlimited. Because judges can set any harm level in $[0, D]$, *D* represents not only the maximum level of harm the injurer can impose on the victim, but also the extent of fact discretion. When *D* is higher, the potential for fact discretion is also greater. Empirically, *D* can be interpreted as measuring the complexity of a dispute: the higher is the number of material dimensions determining harm, the greater is judges' discretion in estimating *h.*

Because judges engage in fact discretion to promote their own agenda, to study fact discretion we need to specify judicial preferences. We assume that judicial preferences are defined over damages, so that judge *j*'s loss from setting damages *d* in case *h* is equal to:

$$L_j = \left| d - d_j^*(h) \right|. \tag{3}$$

$d_j^*(h)$ is the judge's preferred level of damages when true harm is *h*.

We assume that there are three types of judges. A share $\pi/2$ of them are biased in favor of the injurer (*Pro-I* henceforth) and have $d_j^*(h) = 0$. A share $\pi/2$ of them are biased in favor of

the victim (*Pro-V*) and have $d_j^*(h) = D$. The remaining $(1-\pi)$ judges are *Unbiased* and have

$d_j^*(h) = h$. Parameter π measures the polarization of judicial preferences: the larger is π, the

greater the proportion of biased judges in the population. Empirically, π may measure the

political or social sensitivity of a dispute. For example, politically sensitive areas of law such as

environmental torts or discrimination disputes are characterized by a higher π[5].

The timing of the model is as follows: at t = 0, *I* observes *h* and takes precautions; at t =

1/2, *V* is injured; at t = 1, a trial judge is randomly selected from the population of judges. The

selected judge observes *h*, finds *h'* that is potentially different from *h*, and awards damages $d(h')$

to the victim. A judge's fact finding policy is thus summarized by the function $h'(h)$ assigning to

every true harm level *h* the utility maximizing harm level $h'(h)$ actually found by the judge.

Because the set of admissible damages from harm depends on the prevailing legal rule $d(h)$, the

rule interacts with judicial bias in shaping adjudication.

We focus on the strict liability regime but distinguish two types of prevailing legal rules

within that regime. The first, "settled law," is defined as $d(h) = h$ for all possible kinds or levels

of harm.[6] This definition of settled law includes both strict liability *and* the assumption that all

harms are legally cognizable – the situation that yields first best precautions under standard

assumptions. The second legal rule we consider, "unsettled law," refers to the situation in which

not all fact situations have been previously considered by courts, and therefore the function

$d(h)$ has been defined only for some fact situations *h*. In the explosives example, it might not

have been settled by precedent whether mental anguish is a legally cognizable form of harm.

Unsettled law tends to be the standard situation in legal cases with factual complexity (Llewellyn

1960, Stone 1985).

---

[5] An alternative possibility is fact-finders' systematic bias, such as pro-government sentiment of an average judge, or jury nullification of unpopular laws. We discuss this matter briefly in Section 6.

[6] Studying fact discretion under negligence rules would complicate the analysis because in that case precautions may jump rather then change smoothly with judicial error. We leave the study of negligence rules for future research.

3. Enforcement of Settled Law under Judicial Fact Discretion

Under settled law, if judges find true harm $h'=h$, they correctly set damages $d(h)=h$, and the first best is attained. However, under fact discretion, judges may fail to find true harm (and damages), and the first best may not be attained. We study the impact of fact discretion under settled law by focusing on its effect on precautions, welfare, but also – from an empirical standpoint – on the patterns of accidents and the predictability of adjudication.

Consider how trial courts enforce settled law. At any harm level h, *Pro-I* judges find $h'(h)=0$ so as to set their preferred $d=0$; *Pro-V* judges find $h'(h)=D$ so as to set their preferred $d=D$; Unbiased judges find the truth $h'(h)=h$ so as to set their preferred $d(h)=h$, which is also efficient. It is then easy to see how fact discretion affects precautions. Since the injurer chooses precautions before he knows the judge's type, under settled law his optimal precautions $\tilde{p}_{sl}(h)$ at every level of harm h are equal to expected damages, $E(d(h')|h)$:

$$\tilde{p}_{sl}(h) = E(d(h')|h) = \pi D / 2 + (1-\pi)h, \tag{5}$$

which is the average of the damages set by *Pro-V*, *Pro-I*, and Unbiased judges ($d=h$).
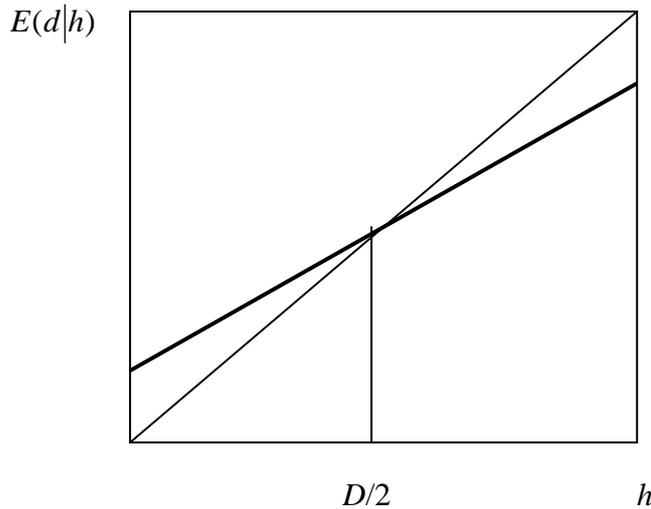

$E(d|h)$

$D/2$ h

Figure 1.

Figure 1 plots the expected damages with and without fact discretion as a function of true harm: the bold line plots *I*'s precautions under fact discretion (for $\pi > 0$), the diagonal represents

10

the first best level of damages obtained under no fact discretion. Under judicial fact discretion, the level of precautions is generally not optimal (except at *h=D/2*). When $h < D/2$, expected damages are too high and the injurer takes excessive precautions. When $h > D/2$, expected damages are too low and the injurer takes insufficient precautions. Under fact discretion, the function linking expected damages to harm is too flat at every *h*. Because biased judges waste information about *h*, expected damages vary too little with harm. In the extreme case where all judges are biased ($\pi = 1$), expected damages (and thus precautions) are constant at $D/2$.

The insensitivity of damages to true harm is intuitive. When harm is close to zero, *Pro-I* judges are close to efficiency. However, *Pro-V* judges find harm far in excess of reality, and so *I* takes excessive precautions. In contrast, when harm is close to *D*, *Pro-V* judges are close to efficiency, but the damages set by *Pro-I* judges are far too low and so *I* takes too few precautions. That is, when harm is low, over-estimation of damages is the most relevant error and the harshness of *Pro-V* judges induces over-precautions. When instead harm is high, damages tend to be under-estimated, so the leniency of *Pro-I* judges induces under-precautions.[7] Because of these distortions, damages cannot vary as much as harm does, implying that the *average* slope of damages with respect to harm is too flat. As Figure 1 shows, in our model of judicial bias this effect is even stronger: damages are too flat at every *h*, not only on average.

These results are related to the research on accuracy in adjudication (e.g., Kaplow 1994), which studies the impact of inaccurate assessment of damages (or liability) on precautions and welfare. Kaplow and Shavell (1996) show that inaccurate determination of damages, as long as it is on average unbiased, does not affect precautions. We show instead that when deviations from accuracy result from judicial bias rather than from random error, it is unlikely that opposite errors cancel out on average because the effect of opposite biases is asymmetric, depending on the level

---

[7]This result does not rely on *h* having bounded support (indeed, increasing the support length *D* strengthens our result). It hinges on the reasonable assumption, implicit in our model, that a judge's return from fact discretion decreases as the true facts become more favorable to his preferred party.

of harm.[8] In this respect, our results relate to Craswell and Calfee (1986), who show that legal error in determining liability (not damages) may exert two opposite effects on precautions: false liability induces over-precautions, false non-liability induces under-precautions. Our model shows that if errors are a product of judicial bias, then either effect may prevail depending on harm. When harm is low the bias of *Pro-V* judges is more important, the first effect dominates, and *I* is overly cautious; when harm is high the opposite is true.

This discussion shows that modeling legal errors as a product of the deliberate decisions of utility maximizing judges rather than purely random actually matters for the evaluation of accuracy in adjudication. In addition, our model shows how, by affecting legal error, judicial preferences π and legal complexity D determine the extent to which precautions deviate from the first best. This can be readily seen by looking at social welfare:

Proposition 1: Under settled law, the loss of social welfare relative to the first best arising from fact discretion is equal to $(\pi D)^2 / 24$. Increases in both π and D reduce welfare; the marginal social cost of D increases in π.

Consistent with our previous analysis, judicial bias is responsible for the cost of fact discretion. When judges are unbiased, they enforce the optimal level of damages irrespective of their ability to misrepresent harm, and the first best is attained. If instead some judges are biased (i.e., if $\pi > 0$), injurers no longer have the incentive to take efficient precautions and the first best is no longer attained. The extent to which precautions (and welfare) deviate from the first best depends on π and D. When more of the judges are biased, expected damages are less sensitive to true harm, over- and under-precautions become more severe, and welfare falls. Likewise, an increase in D raises under- and over-precautions and reduces welfare by increasing the ability of

---

[8]This result holds also if *I* sets *p* by only knowing that harm belongs to a neighbourhood of *h*, provided that such neighbourhood does not coincide with the range of fact discretion $[0, D]$. In the latter case, as in Kaplow and Shavell (1996), damages and precautions are equal to the socially optimal level *D/2*, because judges are on average unbiased.

biased judges to distort damages. Furthermore, as the range of fact discretion *D* widens, biased judges obtain more opportunities to distort the setting of damages, so judicial polarization has a detrimental impact on precautions. Judicial bias is particularly costly in those complex areas of law where biased judges can distort fact finding to the point of undermining the performance of an otherwise optimal legal rule, such as strict liability.

Aside from the welfare costs of fact discretion, what might be some of the *observed* consequences of this behavior of trial judges? First, the analysis has implications for statistical predictability of judicial decisions *from case facts*. By "unpredictability" we mean the extent to which the level of damages cannot be predicted by true harm *h*. Define unpredictability by:

$$\text{unpredictability} = 1 - \rho(d, h), \tag{4}$$

where $\rho(d, h)$ is the correlation between damages and true harm. In the first best, $\rho(d_{fb}, h) = 1$. We thus measure the unpredictability of adjudication by the shortfall of the correlation between damages and harm from its first best level. We then have:

Corollary 1: Under settled law, $\rho(d, h) = (1 - \pi) / \sqrt{2\pi + 1}$. Unpredictability increases in π.

Not only does fact discretion cause unpredictability, but the latter is more severe when more of the judges are biased. Thus, even when legal rules are fixed, via fact discretion judicial bias introduces volatility into adjudication, and the more so the greater the dispersion of judicial views. For example, we expect less predictability from facts in politically sensitive cases. However, in this model, unpredictability is unaffected by the extent of fact discretion *D*: a larger *D* lets biased judges distort fact finding more, but it also allows unbiased judges to optimally set damages over a broader range.

Second, the analysis suggests that while fact discretion makes facts less helpful in predicting trial outcomes, it makes judicial preferences more helpful for doing so. If we have independent measures of judicial bias, these measures will predict resolution of identical disputes.

There is by now an enormous literature indicating that race, gender, and the party of the nominating President affects the decisions of appellate judges, especially in politically sensitive cases. Some of the key studies are George and Epstein (1992), Brenner and Spaeth (1995), Revesz (1997), Pinello (1999), Klein (2002), Sunstein, Schkade and Ellman (2004), and Hansford and Spriggs (2006). Unfortunately, we do not know of similar studies for trial courts.

Third, the results on precautions suggest that, in the areas of law where judicial fact discretion is pronounced, we should see a large number of severe accidents as injurers take most insufficient precautions precisely in those situations. Indeed, in our model, fact discretion does not affect the total number of accidents (which is equal to D/2, as in the first best), but increases the relative likelihood of accidents with higher h.

4. Fact Discretion and Judicial Review

A second possible determinant of fact discretion – pertinent to judges but not juries -- is judicial review. Our model of judicial review relies on the generally accepted idea that appellate courts take the trial courts' fact finding as given (except in the cases of "clear error"), but can reverse trial courts if the law was misapplied to the found facts. Appellate courts resolve conflicts over law, not those over facts. The acceptance of trial court's fact finding by appellate courts is a central feature of common law, which distinguishes it from the civil law tradition. One explanation is the greater reliance of common law adjudication on open trials and on oral examination of witnesses at trial as a strategy of gathering evidence, which is not easily compatible with appellate review of fact finding (see Merryman 1985, Glaeser and Shleifer 2002).

For concreteness, suppose that a (randomly selected) trial court solves dispute h by choosing a vector *(d',h'(h))*, where $h'(h)\in[0,D]$ represents the trial court's (potentially distorted) finding of facts and $d'\in[0,D]$ represents the corresponding level of damages set by the judge. After the trial, the case is automatically appealed. The appellate court can either affirm or

reverse the trial court's ruling. We assume that trial judges dislike being reversed and incur a psychological or reputational loss $r > 0$ when this happens. We model the appeals process as follows. An appeals court is randomly selected from the population of such courts. The preferences of appeals courts are again given by (3) and are distributed identically to those of the trial courts. The selected appeals court verifies whether the facts $h'(h)$ found by the trial court warrant the applied level of damages d'. If they do, the appellate court affirms the trial court's ruling. Otherwise, the appellate court reverses the ruling.

In deciding whether to affirm or to reverse, the appellate court maximizes its utility but is compelled to apply the prevailing legal rule. If – given the trial court's fact finding $h'(h)$ – the trial court's damages d' are consistent with the prevailing legal rule, then the appellate court must affirm, even if its bias tempts it to set a different level of damages. In contrast, if the law specifies that a level of damages $d'' \neq d'$ should be set at $h'(h)$, then – irrespective of its preferences – the appellate court must reverse and award d''. The more interesting case arises when the law is unsettled, in that for some facts $h'(h)$ the prevailing legal rule does not specify what level of damages should be correctly applied. In the explosion example, suppose that precedents have not settled whether victims should be compensated for mental suffering. As we show in Section 4.2, in that case the decision to affirm or to reverse crucially depends on the appellate court's bias.

4.1 Judicial Review under Settled Law

An immediate consequence of the working of judicial review in our model is that, under settled law, trial courts can avoid reversal and still be able to set their preferred damages by simply distorting the facts. For example, if harm is h_0 but the trial judge wants to set $d' = h_1 \neq h_0$, then he just needs to find $h'(h_0) = h_1$. Because the appellate court must take facts h_1 as given, it cannot reverse damages $d' = h_1$, which are precisely those mandated by strict liability for the facts found. Reversal would only occur if the trial court finds h_0 but sets $d = h_1$ since then

15

$d = h_1$ is a misapplication of the law to the facts. Yet the trial court never chooses to follow the latter strategy: by exploiting fact discretion, it can set $d = h_1$ and avoid reversal.

Under fact discretion, then, judicial review is irrelevant when the legal rule is strict liability and all harm is legally cognizable. The flexibility of strict liability allows biased trial courts to distort the facts and to set their desired outcomes, and unbiased trial courts to find true harm and to set the efficient damages. As a result, introducing judicial review does not affect unpredictability and welfare. We are back to the findings of Section 3.

4.2 Judicial Review under Unsettled Law

Trial courts often deal with cases in which the mapping from true harm to damages remains unsettled by previous legal rulings. Because such gaps in the law are filled by appellate courts, a trial judge's freedom to set damages is limited by the appellate review of his decision.

This situation, which we call unsettled law, is typical in common law, where legal rules are a by-product of judges resolving specific disputes. When existing precedents fail to exhaust all factual circumstances, and new facts arise in a case, a trial judge who reports these facts truthfully must consider which precedent is controlling. After he renders his decision, the losing party may appeal his ruling by insisting that a more favorable precedent should be applied to the facts found by the trial court. An appellate court must then decide whether, given these facts, the current case as a matter of law is "closer" to the plaintiff's or the defendant's preferred precedent. The trial court can avoid this appellate scrutiny, and possible reversal, by simply distorting the facts of the case so that one of the potentially governing precedents applies exactly. When the law is unsettled and the mapping between new facts and damages requires interpretation, trial judges' incentives to engage in fact discretion depend not only on their biases, but also on the preferences of appellate courts and the cost of reversal. In this situation, fact discretion is no longer an exclusive prerogative of biased judges.

We capture the idea of unsettled law by studying the case with two precedents governing damages in the tort between *I* and *V*: one of them is the case (*h=0, d=0*); the other is the case (*h=D, d=D*). [9] For harm levels falling outside the existing precedents, i.e. for $h \in (0, D)$, the law is silent. This situation is represented below, with the two precedents highlighted in bold.

$$0 \qquad\qquad\qquad\qquad\qquad\qquad D$$

Figure 2.

To choose damages, an appellate court interprets the current case in light of existing precedents. It may deem *h'* sufficiently analogous to *h=0* and resolve the legal ambiguity in favor of *d=0*. Alternatively, the appellate court may hold that *h'* falls within the realm of the *d=D* precedent and set damages accordingly. Finally, the appellate court may decide that *h* is sufficiently different from both precedents and award a third (new) level of damages. In this latter case, the appellate court distinguishes *h* from existing precedents and fills a gap in the law.

As an example, suppose that to properly estimate the total harm $h \in (0, D)$ a new influence on harm must be introduced (e.g., mental anguish in an explosion), which was not considered in existing precedents. Suppose further that the trial court correctly finds this new aspect of harm. An appellate court may deem the new aspect of harm immaterial and set damages by analogy with existing precedents, or it may accept the new influence into adjudication, distinguish the current case from precedents, and deliver a new ruling.

The choice of different appeals courts among these alternatives is key to understanding how judicial review affects trial judges' incentives to engage in fact discretion. To avoid reversal, a trial court considering a report of true facts $h \in (0, D)$ (i.e., considering whether to introduce the new influence on harm) must also consider what level of damages the appellate courts will itself

---

[9] To avoid the implication that, if the law is unsettled, very few damages are awarded in equilibrium, one could allow trial courts some range of discretion in setting damages around the precedents' reference points. In this case none of our results would substantially change.

choose at $h \in (0, D)$. When all appellate courts are unbiased, the matter is simple. The trial court simply reports the true facts and sets $d = h$, and the appellate court affirms by distinguishing the case from the precedents. In this case, unsettled law does not matter[10].

But matters differ when some appellate courts are biased. Consider appellate courts' reaction to a generic trial ruling *(d',h'(h)).*

Lemma 1: At *h'=0* and *h'=D* the trial court is affirmed if and only if *d'=0* and *d'=D*, respectively. If *h'* $\in$ *(0,D)*, *Pro-V* appeals courts reverse any *d'<D*, *Pro-I* appeals courts reverse any *d'>0*, while *Unbiased* appeals courts reverse any $d' \neq h'$.

Not surprisingly, appellate courts exploit legal ambiguities to affirm their biases. If the facts squarely fit into existing precedents (i.e., *h'=0* or *D*), there is no legal ambiguity and appellate courts affirm trial court rulings, consistent with those precedents (i.e., *d'=0* or *D*, respectively). But if $h' \in (0, D)$, the resolution of legal uncertainty over damages depends on the bias of the appellate court reviewing the case. A trial court's finding of no mental anguish is radically different from the finding that mental anguish should be excluded from the damage calculation. The former decision simply cannot be reversed, but the latter one can. Indeed, suppose the only harm at stake is mental anguish *h*, and the trial court rules that it is not cognizable for the damage calculation, so *d=0.* In this case, a *Pro-I* appellate court affirms the ruling that mental anguish is not cognizable, a *Pro-V* reverses and sets *d=D*, by analogy with the existing precedent of severe harm. An unbiased appellate court also reverses, rules that mental anguish is an admissible harm, but sets damages *d = h*. When existing precedents can be interpreted differently, the decision to reverse or to affirm a trial court's ruling depends on the appellate court's bias. A trial judge averse to being reversed might opt for distorting the facts rather than subjecting himself to the uncertainty of legal oversight (cf. *Garratt v. Dailey*).

---

[10] This benign view of unbiased appellate review is consistent with Bueno de Mesquita (2002) and Shavell (2005).

How does judicial review affect fact discretion, as represented by trial courts' fact finding policy $h'(h)$? Consider the case of an unbiased trial judge. Lemma 1 implies that, to avoid reversal, the trial judge must fit the case within existing precedents. Suppose that $h \in (0, D)$. If the trial judge engages in fact discretion and rules $(d' = 0, h' = 0)$, he loses $|0 - h| = h$; if he rules $(d' = D, h' = D)$, he loses $|D - h| = D - h$. In neither case he gets reversed. If instead the trial judge finds $h'(h) \in (0, D)$, his optimal strategy (by Lemma 1) is obviously to find the truth and rule $(d' = h, h' = h)$. In this case, his expected loss is:

$$(\pi / 2)(h + r) + (1 - \pi)0 + (\pi / 2)(D - h + r) = \pi(r + D / 2) \qquad (6)$$

The first term is the trial judge's loss from reversal by a *Pro-I* appeals court that sets $d'=0$. The third term is the trial judges' loss from reversal by a *Pro-V* appeals court that sets $d'=D$. If the appeals court is unbiased, it affirms the unbiased trial judge's ruling, who then loses nothing.

By comparing trial courts' losses from alternative strategies, we see that the unbiased judge's fact finding policy trades off the gain from setting first best damages against the total reversal cost. Reversal is costly to the trial judge for two reasons. First, the appellate court may set damages too far away from the trial judge's preferred ones. Second, the trial judge bears the reputational cost r. These reversal costs are higher when the share of biased appellate courts (π) is higher and when the reputational cost r is higher. The trial court rules as follows:

Proposition 2: *Pro-I* trial judges set $(d'=0, h'(h)=0)$. *Pro-V* judges set $(d'=D, h'(h)=D)$. *Unbiased* judges rule as follows: there are two thresholds $\bar{h}, \underline{h}$ $(0 \le \underline{h} \le D/2 \le \bar{h} \le D)$ such that they set $(d'=0, h'(h)=0)$ for $h < \underline{h}$, $(d'=D, h'(h)=D)$ for $h > \bar{h}$ and $(d'=h, h'(h)=h)$ otherwise.

Because biased judges indulge their biases, they engage in fact discretion, but also try to minimize the chance of reversal. When existing precedents include biased judges' preferred points $d=0$ and $d=D$, these judges can attain their individual optima with no fear of reversal.

In contrast with our previous results, Proposition 2 shows that, under judicial review, even unbiased judges engage in fact discretion. Unbiased judges would ideally avoid fact discretion and set (*d'=h,h'(h)=h*). However, with unsettled law, fear of reversal on appeal encourages them to distort fact finding. This is due to the risk that the appellate court is biased. Because unbiased appeals courts never reverse (*d'=h,h'(h)=h*), unbiased trial courts facing no risk of appellate bias could always find $h'(h) = h$, confident that their ruling will be affirmed. Figure 3 depicts adjudication by unbiased trial courts in terms of the cutoff points $\underline{h}$, $\overline{h}$ for various decisions.

$$(d'= 0, h'(h) = 0) \quad (d'= h, h'(h) = h) \quad (d'= D, h'(h) = D)$$

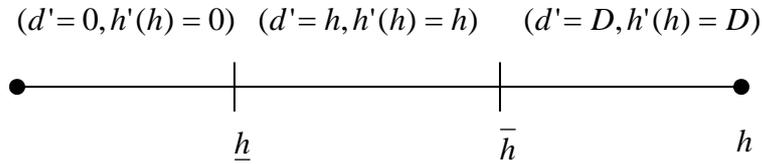$$\underline{h} \qquad\qquad \overline{h} \qquad\qquad h$$

Figure 3.

Notice that $\underline{h} = \min[\pi(r + D/2), D/2]$ and $\overline{h} = \max[D - \pi(r + D/2), D/2]$. Hence, greater polarization of appellate courts' views (π) and greater reputational loss *r* render reversal more costly, expanding the unbiased trial courts' exercise of fact discretion. In particular, if $\pi > 1/(2r + D)$, then $\overline{h} = \underline{h} = D/2$. Now the cost of reversal is so high that trial courts do not benefit from distinguishing and even unbiased judges *always* engage in fact discretion.

What is the impact of unbiased courts' fact discretion on precautions and welfare when the law in unsettled? We answer this question by focusing on the case where $\pi > 1/(2r + D)$.

Figure 4 shows the impact of unsettled law on damages and precautions when $\pi > 1/(2r + D)$. Besides being analytically more tractable, this case allows a sharper evaluation of the impact of unbiased judges' exercise of fact discretion. Compared to settled law, where expected damages smoothly increase with harm, the key feature of unsettled law is that damages experience a sharp jump at *h = D/2*. Indeed, because under unsettled law unbiased judges fear

reversal, they engage in fact discretion, finding *h'*=0 for $h \leq D/2$, and *h'*=*D* otherwise. Now

damages are a step function and the only information used in adjudication is whether *h* is above or

below *D/2*. As before, damages are on average too flat, as the injurer takes excessive precautions
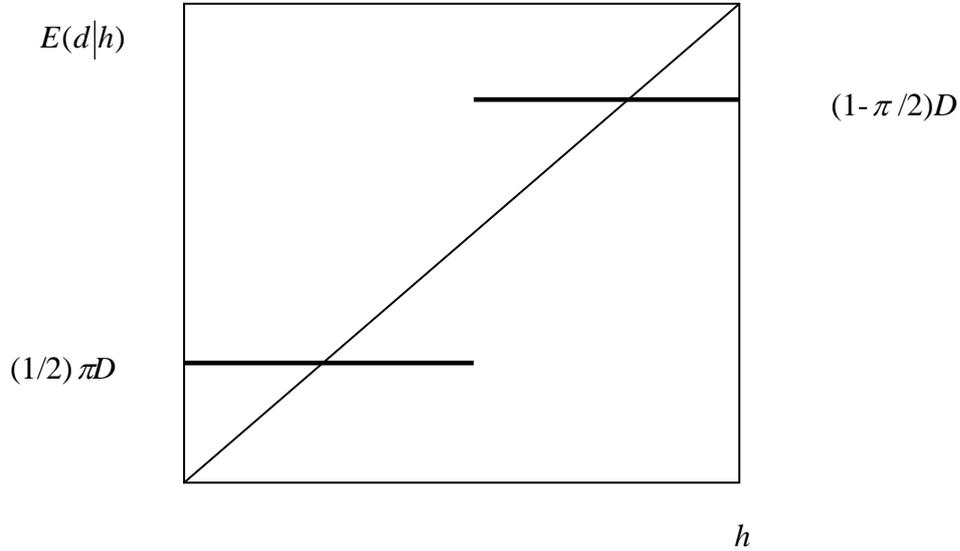
when harm is low and insufficient ones when harm is high.

$E(d|h)$

$(1-\pi/2)D$

$(1/2)\pi D$

h

Figure 4.

For the impact of fact discretion on welfare, we find:

Proposition 3: Under unsettled law, if $\pi > 1/(2r+D)$, the loss of welfare relative to the first

best is $\left[1+3\pi^2-3\pi\right]D^2/24$. This loss of welfare decreases in π for $\pi \leq 1/2$, and increases in π

otherwise. Judicial polarization has a stronger impact on welfare when *D* is larger.

With unsettled law, fact discretion lowers social welfare relative to the first best.

Furthermore, when $\pi \geq 1/2$, greater polarization reduces social welfare. As with settled law,

under unsettled law judicial bias reduces the extent to which damages vary with harm, thereby

inducing over and under-precautions. Yet, in contrast to settled law, under unsettled law greater

polarization may be beneficial. We fully discuss this non-monotonicity in Section 6. The basic

intuition is that an increase in the share of biased judges reduces the jump in damages and thus in

precautions at *h = D/2*. Because the marginal cost of precautions is increasing, this reduction may

in turn beneficially reduce the average cost of precautions by smoothing them across harm levels. As before, the extent of fact discretion D magnifies the (positive or negative) impact of judicial bias on social welfare. Consistent with Proposition 1, bias is especially important in complex areas of law, where the scope for fact discretion is greater.

The main observable implications of fact discretion arising from judicial review line up with those arising from judicial biases. It is still the case that measurable judicial biases affect judicial decisions, although now even unbiased judges make biased decisions. It is also still the case that there are too many bad accidents relative to the first best. Finally, it is still the case that judicial decisions are unpredictable from the facts of a case. Specifically, we have:

Corollary 2: Suppose that $\pi > 1/(2r + D)$. Then, under unsettled law, $\rho(d, h) = (1 - \pi)\sqrt{3/4}$. Unpredictability increases in π.

As before, fact discretion reduces the correlation between damages and harm; predictability decrease in judicial polarization but is unaffected by factual complexity D.

One predicted consequence of judicial fact discretion arising from judicial review is new. Specifically, we expect that in complex and unsettled areas of law, where determination of liability requires answers to a variety of factual questions, the exercise of fact discretion would be more pronounced. If a researcher had an independent ability to observe the facts (perhaps from the documentary record), and compare them to the judge's summary of the evidence, it is precisely in these complex fact-intensive areas of law that we expect the greatest mismatch between the true facts and the judge's representation of those facts. For it is precisely in these areas of law that misrepresentation of the evidence best protects the judge from reversal.

22

5. Fact Discretion and Litigation

So far we have focused on the behavior of trial judges, and treated the conduct of the litigants mechanically. In this section we focus on the impact of fact discretion on the behavior of litigants. [11] We compare the outcomes obtained under settled law, with those obtained under unsettled law. We show that with fact discretion: a) unsettled law encourages the litigants to take partisan positions during trial as opposed to presenting the truth, and b) litigants may go to trial even when they agree on the facts of the case, because fact discretion introduces extra-factual uncertainty about the trial outcome, especially under unsettled law.

Consider the following game played by the litigants and the judge. After ending up in court, but before the type of the judge is revealed, each party $P = I, V$ sends to the judge a message h_P concerning the level of harm. A litigant's message about harm represents his position in court and conveniently summarizes a possibly multi-dimensional vector of pieces of evidence that the litigant submits to the judge. In this respect, a litigant's position is the more partisan, the closer it is to his desired level of damages. If the parties' messages are identical, the judge takes them as the true harm. If instead the messages diverge, the judge decides whether to accept one of them or to find out the true level of harm. We thus relax the assumption that judges find out h and instead assume that, to observe true harm, the judge must incur a search cost of $k|h_I - h_V|$, with $0<k<1/2$. This assumption captures the idea that, when the parties' messages are far apart, their positions are barely informative and the judge must collect more data on his own and bear a higher cost. This assumption greatly simplifies the analysis, but is not crucial: our main results also hold with a fixed search cost. The judge exits this stage with a conclusion about harm $\hat{h}$ (obtained from one or both parties or found independently). At this point, the judge

---

[11] For simplicity, we abstract from I's ex-ante choice of precautions and only study the choice of litigation vs. settlement, but one could extend the analysis to incorporate precautions.

issues his preferred ruling $(d', h'(\hat{h}))$ based on $\hat{h}$.[12] We capture competition among litigants by assuming that – if the judge observes true harm – he negligibly shades damages away from his preferred level toward the less partisan litigant.[13]

5.1. Fact Discretion and Litigation under Settled Law

The Litigants' Positions under Settled Law

Under settled law, biased judges are irrelevant for determining litigants' messages because these judges are uninterested in finding out true harm. From any two messages (h_I, h_V) sent by litigants, a biased judge uses only the pieces of evidence pointing to his preferred outcome: *Pro-V* judges rule $(d'=D, h'=D)$, *Pro-I* judges rule $(d'=0, h'=0)$. Only the fact-finding strategy of unbiased judges shapes the litigants' positions in court.

Consider now an unbiased judge's fact-finding. Suppose that litigants' messages are h_I, h_V, with $h_I \le h \le h_v$, where h is true harm. (This is always the case in equilibrium.) Then, if the judge decides to find out the truth, he spends $k(h_V - h_I)$ and rules $(d'=h, h'=h)$, bearing a loss of $k(h_V - h_I)$. If instead the judge does not find out h, his loss is identical if he sets either $d = h_I$ or $d = h_V$ and is equal to $(h_V - h_I)/2$. Because $k<1/2$, the judge's loss from finding h is always smaller than that from not finding it. As a consequence, for any two messages h_I, h_V the judge always finds the truth under settled law. What is the impact of such judicial strategy on the parties' optimal choice of h_I, h_V?

Proposition 4: Under settled law, litigants set $h_I = h_V = h$ for any h. Unbiased judges rule $(d'=h, h'(h)=h)$, *Pro-I* judges rule $(d'=0, h'=0)$, *Pro-V* judges rule $(d'=D, h'=D)$.

---

[12] Even if the judge finds out true harm (i.e., $\hat{h}=h$), we still allow him to flexibly distort fact finding and find a different level of harm. Disallowing this flexibility would strengthen our results.
[13] Including in the model a cost for the judge of deviating substantially from the reports of both litigants would create an even stronger punishment for partisanship and reinforce our results.

By allowing unbiased judges to accurately fine tune damages to harm, strict liability with all damages being legally cognizable gives unbiased judges a strong incentive to verify harm. As a result, each litigant has an incentive to move closer and closer to the actual h so as to avoid having the judge shade damages against him. This type of settled law dampens the partisanship of the parties by giving unbiased judges a strong incentive to scrutinize the litigants' positions in court. When $k > 1/2$, judges never decide to find out the truth and Proposition 4 no longer holds. We assume $k \leq 1/2$ because this case best illustrates the difference between settled and unsettled law in shaping litigants' partisanship.

Litigation vs. Settlement under Settled Law

What is the impact of fact discretion for the incidence of litigation under settled law? If settlement is cheaper than litigation, then parties litigate only in the presence of bargaining frictions. Such frictions may result from litigants' over-optimism about the chance of winning the trial (Landes 1971, Posner 1972) or from their different (private) information about the merits of the case (Bebchuk 1984). In our simple model with parties agreeing on the facts of the case, no litigation arises in the absence of fact discretion, because parties are perfectly informed about harm and courts predictably set damages equal to the correct level of harm. However, litigants may still have heterogeneous beliefs or information precisely because of fact discretion, since fact discretion introduces extra-factual uncertainty about the outcome of the trial. In this precise sense, fact discretion promotes socially wasteful litigation.

In particular, under settled law, damages do not solely depend on harm, but also on judicial bias. Disagreement over judicial bias may thus lead to litigation if each party is optimistic about the likelihood of getting a favorable judge. This latter scenario is even more plausible when, as argued by Frank (1930), a judge's bias reflects his idiosyncratic sympathy or antipathy toward specific litigants rather than more stable policy preferences. For example, a

judge may be annoyed with a lawyer from an earlier case, sympathetic to one who previously clerked for him, or differential to a government attorney who works in the same building.

To see how in this case litigation may arise in our model, consider the parties' decision to settle or litigate in case *h*. For simplicity, we follow Yildiz (2004) and study the case where the parties' failure to settle is due to their heterogeneous beliefs rather than to asymmetric information. Suppose that the injurer believes that the share of Pro-*I* judges is inflated by a factor $(1+\delta)$ and that of Pro-*V* judges is deflated by a factor $(1-\delta)$, while the victim misperceives the share of Pro-*I* and Pro-*V* judges the other way around. $\delta \geq 0$ captures the divergence in litigants' beliefs: when δ is higher, both parties are more optimistic of being tried by a favorable judge. If the individual litigation cost is *c*>0, litigants' expected payoffs from litigating case *h* are:

$$Eu_{injurer} = -(\pi/2)(1+\delta)0 - (1-\pi)h - (\pi/2)(1-\delta)D - c$$
$$Eu_{victim} = (\pi/2)(1-\delta)0 + (1-\pi)h + (\pi/2)(1+\delta)D - c$$

(7)

With these payoffs, the parties fail to find a mutually profitable settlement amount paid by the injurer to the victim and thus litigate if and only if:

$$(\pi/2)\delta D \geq c$$

(8)

From this expression it immediately follows that:

Proposition 5: If $D = 0$, the parties always settle. If $D > 0$, then there exists a $\underline{\delta} \in [0,1]$ such that the parties litigate if and only if $\delta > \underline{\delta}$. $\underline{\delta}$ decreases in *D* and π.

With no fact discretion, only harm matters for setting damages, not judicial bias. As a consequence, the parties settle when *h* can be verified. In contrast, because under fact discretion judicial bias affects the setting of damages, when $D > 0$ the parties litigate if they are sufficiently optimistic about the chance of getting a favorable judge (i.e., if δ is high enough). Importantly, the parties' optimism and fact discretion interact: the extent to which divergence in beliefs leads to litigation increases in the scope for fact discretion. When this scope is great, biased judges

change the actual value of damages by a lot. As a result, the stakes of the case (the payoff difference between winning and losing) increase and magnify the impact of the heterogeneity of beliefs δ. By introducing extrinsic factors such as judicial bias into trials, judicial fact discretion may lead to wasteful litigation.

This analysis of litigation under fact discretion yields two empirical predictions. First, litigation should be more prevalent in the politically or socially charged areas of law, where judicial views are more likely to be polarized. Likewise, litigation should be more prevalent in more complex areas of law, where the application of legal rules requires the verification of many factual issues, even if these legal rules are clear and unambiguous.

Second, when parties hold similar beliefs on the distribution of judicial bias and litigation does not occur, we predict that – due to fact discretion – pre-trial (or more precisely, pre-revelation of judge's type) settlement amounts in different cases would cluster around the mean settlement, especially if judicial polarization is high. This finding stands in contrast with the standard prediction of Priest and Klein (1984) that pre-trial settlements are especially likely to occur when the facts of a dispute are clear. In their model, settlement amounts should reflect the disparate facts of individual cases and presumably display considerable variance rather than converge to the mean. On the other hand, like Priest and Klein our model predicts that settlement amounts should spread out once the identity of the judge, and therefore presumably his type, is revealed. In such settlements, the party whose position the judge is expected to be sympathetic to should receive most of the benefit in the settlement.

5.2. Fact Discretion and Litigation under Unsettled Law

The Litigants' Positions under Unsettled Law

As we showed in Section 4, when the law is unsettled, even unbiased judges may set damages at the extremes. This fact has two key implications. First, unbiased judges may prefer

to remain uninformed, because – in contrast to settled law – unsettled law does not allow them to use much information anyway. Second, litigants may take partisan positions to cater to judges' desire to fit the case into the existing precedents.

To see how this works, suppose that $\pi > 1/(2r + D)$. Recall that, in this case, unbiased judges only consider whether harm is larger or smaller than $D/2$ to choose between $d=0$ and $d=D$. This fact has important consequences for unbiased judges' incentive to find out the true h. Indeed, suppose that under the law [$d(0)=0$, $d(D)=D$], parties send $h_I = 0$ and $h_V = D$. Then, if the judge finds out true harm, his expected loss is:

$$\int_0^{D/2}(h/D)dh + \int_{D/2}^{D}\left[(D-h)/D\right]Ddh + kD = D\left(\frac{1}{8}+k\right) \tag{9}$$

This expected loss equals the judge's average loss from setting $d=0$ when $h \leq D/2$ and $d=D$ when $h > D/2$, plus the search cost kD.

If instead the judge does not find out harm, his expected loss is the same if he sets $d=0$ and $d=D$ and is equal to $D/2$. But then, if $k>3/8$, the judge prefers not to find the true level of harm h. This preliminary finding indicates that, under unsettled law, the judge has less of an incentive to find the true h than under settled law. When $k \in (3/8, 1/2)$, the judge finds the true h under settled but not under unsettled law, irrespective of (h_I, h_V). The judge's gain from acquiring information is smaller under unsettled law because, fearful of reversal, he must choose between $d=0$ and $d=D$ and thus cannot use all the information acquired by setting $d=h$ as under settled law. Because the judge remains uninformed about the true h, not only does he choose either $d=0$ or $d=D$, but he is also indifferent between these two outcomes. These observations imply:

Proposition 6: Suppose that $k>3/8$. Then, under unsettled law, the parties' messages are $h_I = 0$ and $h_V = D$. *Pro-I* judges rule ($d'=0, h'=0$), *Pro-V* judges rule ($d'=D, h'=D$), while Unbiased judges remain uninformed and randomize among these two outcomes.

Because with unsettled law the judge remains uninformed and wants to fit the case into an existing precedent, competition between parties in picking their messages is radically different from that prevailing under settled law. Now competition leads to extreme partisanship, not to convergence to the truth. Indeed, because it is never optimal for the judge to find out the true level of harm, he is indifferent between extreme versions of the facts. Furthermore, to avoid reversal, the judge always endorses a partisan message such as $h_I = 0$ or $h_V = D$ as opposed to a message claiming that h is in the middle.[14] As a result, litigants compete by proposing extreme views so as to cater to the judge's demand for precedent-fitting narratives that render reversal less likely. In court, plaintiffs overreach and over-claim, while defendants refuse to acknowledge even the slightest liability for harm, each hoping that the judge simply buys their story[15].

Litigation vs. Settlement under Unsettled Law

The previous analysis helps shed light on the litigants' decision to go to trial under unsettled law. In Section 5.1 we saw that, under settled law, one determinant of such decision was over-optimism about judicial bias. We now show that under unsettled law litigation is more likely, as disagreement over the judge's bias is not even necessary to obtain litigation.

Key to this finding is the result (Proposition 6) that, when even unbiased judges do not set damages in the middle, they deliberately choose to remain uninformed and thus indifferent among extreme outcomes. In such a case, litigants can hope to sway adjudication to their side through courtroom tactics, persuasion techniques, and so on. As a result, litigants' optimism about their ability to sway and influence the decision of an uninformed and indifferent judge can lead them to litigate, irrespective of their beliefs about the distribution of judicial bias.

---

[14] In particular, divergence is the unique equilibrium if it is even negligibly cheaper for the judge to accept one of the harm level presented by the parties than find an entirely new harm level.

[15] One consequence of this equilibrium is that information does not trickle up to appellate courts, which of course slows down its evolution.

To see how this argument works in our model, parameterize the parties' optimism or overconfidence about their ability to sway an unbiased judge under unsettled law with $\sigma > 0$. The injurer (victim) believes that he will be able to influence unbiased judges to set $d=0$ ($d=D$) with probability $(1/2 + \sigma)$. Then, much in the spirit of expression (8), settlement fails when:

$$[(\pi/2)\delta + (1-\pi)\sigma]D \geq c \qquad (10)$$

Just as under settled law, divergence in beliefs as to the proportion of biased judges (δ) in the population fosters litigation. However, under unsettled law there is an additional factor promoting litigation, namely the litigants' optimism (σ) concerning their ability to move an unbiased but uninformed judge to their side. Once more, this effect is stronger the larger is D. Under settled law, the impact of σ is downplayed because unbiased judges have a stronger incentive to become informed. Furthermore, while under settled law the parties readily settle after knowing the judge's type, under unsettled law they may fail to do so even if the judge is unbiased because disagreement remains until the ruling is released. As a result, fact discretion promotes litigation to a greater extent when the law is unsettled.

To sum up, judicial fact discretion shapes the parties' behavior in court and the likelihood that they litigate rather than settle. Fact discretion boosts litigation by making trial outcomes depend on such extrinsic factors as a judge's beliefs, courtroom tactics, and persuasion, over which the litigants are likely to have heterogeneous beliefs (or asymmetric information). Litigation is most likely to occur in new and unsettled areas of law, where judges distort facts to conform to available precedents. We also find that, in contrast to settled law, unsettled law invites the parties to take extreme positions in court by creating a judicial preference for extreme if inaccurate versions of the facts: judges only care about fitting them into an existing precedent so as to avoid reversal. As a result, the litigants take extreme positions against each other.

Taken together, the results in sections 3-5 suggest that the common law system of dispute resolution will perform particularly poorly when the cases are factually complex, the law is

unsettled, and fact-finder preferences are important for the determination of damages (or for that matter of liability). These conditions seem to describe adequately the determination of damages for pain and suffering, as well as of punitive damages, in product accident cases. Law and economics scholarship has been highly critical of how damages are set in these situations (Viscusi 1988, 1998, Cooter 1988), blaming the randomness of observed outcomes on the lack of clarity in the law, the sentiments of judges and juries, and the actual complexity of finding the correct answer. These conditions are, of course, a recipe for trouble in our model. Interestingly, legal scholars suggest limiting fact-finder discretion through less flexible rules for damages calculation as a solution to the problem (e.g., Cooter 1988). We take up this topic in the next section.

6. Optimal Legal Rules under Fact Discretion

By underscoring the welfare costs of fact discretion, the previous sections lead one to ask what legal rules minimize welfare losses due to fact discretion? Supposing that the law can be settled, how should it be settled? Should all social harms be legally cognizable for the calculation of damages, or should they be limited? How do these optimal choices depend on judicial polarization? In this section, we shed some light on the optimal design of legal rules under fact discretion by examining the costs and benefits of narrowing the set of legally cognizable harms.[16]

To answer these questions, it is useful to look at the expression of social welfare under alternative legal rules. Since legal rules matter by affecting the mapping $p(h)$ from harm to precautions, expected social losses under a given legal rule are equal to:

$$\int_0^D \left[(1-p)h + \frac{p^2}{2} \right] \frac{1}{D} \, dh = \frac{1}{2} \left\{ D + E(p)[E(p) - D] + V(p) - 2\operatorname{cov}(p,h) \right\} \tag{10}$$

---

[16] The study of optimal legal rules under fact discretion is related to the tradeoff between rules and standards discussed, e.g., by Kaplow (1992, 1994). Also related is the discussion of optimal judicial discretion; see Shavell (2005) and Glaeser-Shleifer (2002).

Here $E(p)$ is the average level of precautions over all possible harm levels, $V(p)$ is the variance of such precautions, and $\operatorname{cov}(p, h)$ is the covariance between precautions and harm. Expression (10) yields two useful insights concerning optimal legal rules in the presence of fact discretion. First, the term $E(p)[E(p) - D]$ should be minimized, implying that average precautions should be ideally equal to average harm, $D/2$. Second, the term $V(p) - 2\operatorname{cov}(p, h)$ indicates that an optimal legal rule should minimize the variability of precautions while at the same time incentivize the injurer to take greater precautions when it is more efficient to do so.

Thinking about this tradeoff gives some guidance as to what legal rules may be optimal under fact discretion. The average cost of precautions is minimized with a legal rule inducing judges to enforce a flat damages schedule. Under fact discretion, this is done through a rule mandating a single level of damages (which sets $V(p) = 0$). The trouble with this rule is that it induces the injurer to take inefficient precautions at different levels of harm, i.e., $\operatorname{cov}(p, h) = 0$. In contrast, the incentive to take efficient precautions can be raised through a legal rule prescribing large gaps between the damage levels judges can set. Intuitively, under fact discretion damages tend to be too flat, as biased judges set extreme damages irrespective of harm. Gaps in damages increase the covariance between harm and damages by encouraging unbiased judges to enforce sharp jumps in damages when harm is high. The cost of such jumps in damages is that they increase the variance of precautions.

These observations suggest that some insight into the design of optimal legal rules under fact discretion can be gained by focusing on legal rules of the form $d \in \left[\underline{b}, \overline{b}\right] \cup \left[D - \overline{b}, D - \underline{b}\right]$, where $\underline{b}, \overline{b} \le D/2$. Figure 5 below represents this class of legal rules, where the segments in bold represent the set of admissible damages:
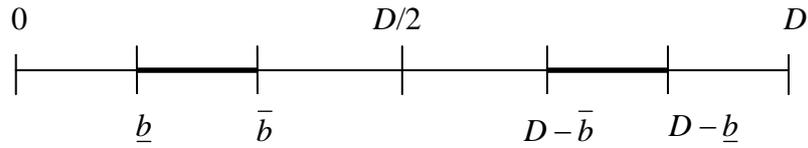
Figure 5.

Although this class of legal rules is not exhaustive, it captures two main aspects of the choice of optimal laws. First, given the symmetry of the set of admissible damages with respect to $D/2$, expected precautions are always equal to average harm $D/2$. Second, the choice of $\bar{b}$ and $\underline{b}$ allows us to flexibly pin down the tradeoff between incentives and variability of precautions. Indeed, an increase in $\underline{b}$ reduces the range of variation for damages, thus reducing $V(p)$ at the expense of incentives, as it reduces $\text{cov}(p,h)$. In turn, a lower $\bar{b}$, by increasing the damage gap between high and low precautions, raises incentives at the expense of a greater variability in precautions by raising the damages jump enforced by unbiased judges. Note that this class of legal rules admits as special cases strict liability with settled law (for $\underline{b}=0,\bar{b}=D/2$), the unsettled law case we studied $d \in \{0,D\}$ (for $\bar{b}=\underline{b}=0$), as well as the rigid rule $d=D/2$ (for $\underline{b}=\bar{b}=D/2$). For simplicity, we abstract from the issues concerning litigation studied in Section 5.[17] Then, solving for the optimal $\underline{b}$ and $\bar{b}$ yields:

Proposition 7: If $\pi < 1/2$, the optimal legal rule is strict liability with all harms being legally cognizable for damage calculations, i.e. $d \in [0,D]$ ($\underline{b}=0,\bar{b}=D/2$); if $\pi \geq 1/2$ the optimal legal rule is to recognize only two levels of harm $d \in \{0,D\}$ ($\bar{b}=\underline{b}=0$).

In the class of rules we consider, the optimal legal rule is either "settled law" as we used the term, or the equivalent of what we called unsettled law, $d \in \{0,D\}$ -- precisely the legal rules

---

[17] The current analysis can be viewed as a special case of our model of litigation obtained when $\delta = 0$, i.e. there is no heterogeneity of beliefs about the distribution of judges biases and $k < 3/8$, i.e. the judges always find the true h.

we considered in the previous sections. In particular, when polarization is low, strict liability with all harms legally cognizable is optimal; when polarization is high, then $d \in \{0, D\}$ is optimal. Intuitively, if the share of biased judges is small, it is best to use the rule achieving the first best in the absence of judicial fact discretion. When instead there are many biased judges in the population, then under that rule damages are too flat and precautions are too inefficient. The legal rule $d \in \{0, D\}$ becomes optimal because it improves incentives by inducing unbiased judges to enforce a very steep damage schedule.

In the current setup, it is never optimal to choose a strict rule eliminating judicial discretion by mandating a unique damage level $d = D/2$. This rule's excessive focus on minimizing $V(p)$ disfavors it relative to $d \in \{0, D\}$. Only if all judges are biased (i.e., $\pi = 1$), then both strict liability and $d \in \{0, D\}$ convey no information about harm, induce I to set $p=D/2$ irrespective of h, and fare just as well as the rigid rule.

The result that the rigid rule $d = D/2$ is never optimal relies on the fact that judges are on average unbiased as there is an equal proportion of *Pro-I* and *Pro-V* judges in the population. In this case, an optimal legal rule should be sufficiently flexible to enhance the incentives of unbiased judges by allowing rulings based on information about harm. A rigid legal rule is optimal, however, when many judges are biased and biased judges tend to systematically favor one party over the other. For example, suppose that all judges are Pro-*I*. Then, under both settled and unsettled law, expected damages are *d=0* and precautions are *p=0*. This outcome is clearly dominated by a rigid rule mandating *d=D/2* for every *h*.

This result is intimately related to the discussion of optimal legal systems when judges are vulnerable to influence by the litigants. Suppose that, instead of having preferences over the outcomes of the dispute, or being concerned with reversal, judges are vulnerable to coercion or bribes, and moreover, one type of the litigant (say, the injurer) is vastly better able to bribe or coerce than the other (say, the victim). This situation is analyzed, for example, by Hay, Shleifer

and Vishny (1996) and Glaeser and Shleifer (2002, 2003). The finding of these papers is very similar to the present one: when judges are systematically biased against a particular group of litigants, social efficiency dictates more rigid legal rules, which minimize judicial discretion.

7. Conclusion.

We have presented two models of judicial fact discretion. In the first, the motivation for the exercise of fact discretion is a trial judge's preference over the outcomes of litigation. This model is probably most relevant for politicized or otherwise emotionally charged disputes. In the second model, the motivation for the exercise of fact discretion is trial judges' aversion to reversal by appellate courts, which leads them to fit the facts of the current dispute into available precedents. This model is probably most relevant for new and developing areas of law, with significant factual complexity and relatively few precedents. For both models, we have shown that fact discretion leads to judicial behavior that is unpredictable from the facts of the case, but predictable from the knowledge of judicial preferences. We have also shown that the exercise of fact discretion leads to systematic distortions in individual behavior, to excessive and acrimonious litigation, as well as to welfare losses.

In conclusion, we briefly mention some issues that are suggested by our model, but that we did not analyze. First, the model implies clearly and perhaps significantly that summaries of relevant facts that accompany written judicial opinions cannot be trusted. As we saw in *Garrett v Dailey* and discussed throughout the paper, when judges summarize the facts, they do so to justify their legal conclusions. If a judge exercises fact discretion, this summary need not reflect the true facts of the case, even as seen and believed by the judge. In some instances, the summary of the facts might be possible to check against other available documents. Unfortunately, from the viewpoint of a researcher or a law student, the judge's summary is often all that is available. This

reality does not necessarily undermine the study of legal principles, but may shed only a dim light on the actual facts of any given case.

Second, we have characterized optimal legal rules, and have shown that the prevalence of biased judges favors rules that limit the exercise of judicial fact discretion. In our specific model, this meant limiting the range of legally cognizable harms. The economic loss doctrine might be one important manifestation of this general principle. There are of course other ways to limit such exercise of fact discretion, such as procedural rules concerning admissibility of evidence or even, as in civil law systems, more extensive appellate review of fact finding. When judicial fact discretion becomes extreme, dispute resolution in court may become socially inefficient. In those instances, adjudication can be replaced by ex ante regulation that does not rely on fact finding.

Third, we have focused our analysis on the exercise of fact discretion by judges, although of course the same phenomenon might be as or more prevalent among juries (Kalven and Zeisel 1966). In the case of juries, legal strategies aiming to control fact discretion tend to focus on the rules of evidence rather than on re-specifications of legal rules that might not impress the juries.

As a final point, we note that this paper is part of a growing body of research that suggests that the consequences and the efficiency of alternative legal arrangements cannot be evaluated without an explicit discussion of preferences and incentives of law enforcers. Rules and arrangements that appear highly desirable with benign law enforcers, such as strict liability with all harms being legally cognizable, lose at least part of their appeal when enforced opportunistically. Judicial fact discretion is but one, although possibly very important, manifestation of this broader problem.

References

Bebchuk, Lucian (1984), "Litigation and Settlement under Imperfect Information," *RAND Journal of Economics* 15, 404-415.

Brenner, Saul, and Harold Spaeth (1995), *Stare Indecisis: The Alteration of Precedent on the U.S. Supreme Court, 1946-1992*, New York: Cambridge University Press.

Brunetti, John (1998), "Searching for Methods of Trial Court Fact-Finding and Decision Making," *Hastings Law Journal* 49, 1491-1509.

Bueno de Mesquita, Ethan, and Mathew Stephenson, 2002, "Informative Precedent and Intrajudicial Communication," *American Political Science Review* 90, 1-12.

Cooter, Robert (1988), "Punitive Damages for Deterrence: When and How Much?," *Alabama Law Review* 40, 1143-1196.

Craswell, Richard, and John Calfee (1986), "Deterrence and Uncertain Legal Standards," *Journal of Law, Economics, and Organization* 2, 279-303.

Dershowitz, Alan, and John Hart Ely (1971), "*Harris v. New York*: Some Anxious Observations on the Candor and Logic of the Emerging Nixon Majority," *Yale Law Journal* 80, 1198-1227.

Frank, Jerome (1930), *Law and the Modern Mind*, New York: Brentano's.

Frank, Jerome (1932), "What Courts Do in Fact," *University of Illinois Law Review* 26, 645.

Frank, Jerome (1951), *Courts on Trial*, Princeton, N.J.: Princeton University Press.

Gennaioli, Nicola (2004), "Contracting in the Shadow of the Law," Mimeo, Harvard University.

Gennaioli Nicola, and Andrei Shleifer (2006), "The Evolution of Precedent," Mimeo.

George, Tracey, and Lee Epstein (1992), "On the Nature of Supreme Court Decision Making," *American Political Science Review* 86, 323-337.

Glaeser, Edward, and Andrei Shleifer (2002), "Legal Origins," *Quarterly Journal of Economics* 117, 1193-1229.

Glaeser, Edward, and Andrei Shleifer (2003), "The Rise of the Regulatory State," J*ournal of Economic Literature* 16, 401-425.

Hansford, Thomas, and James Spriggs, II (2006), *The Politics of Precedent in the U.S. Supreme Court,* Princeton, N.J.: Princeton University Press, forthcoming.

Hay, Jonathan, Andrei Shleifer and Robert Vishny (1996), "Toward a Theory of Legal Reform," *European Economic Review* 40, 559-567.

Henderson, James (2003), "*MacPherson v. Buick Motor Company*: Simplifying the Facts While Reshaping the Law," in Robert Rabin and Stephen Sugarman, eds., *Torts Stories*, New York, NY: Foundation Press.

Kalven, Harry, and Hans Zeisel (1966), *The American Jury*, Chicago, IL: The University of Chicago Press.

Kaplow, Louis (1992), "Rules versus Standards: An Economics Analysis," *Duke Law Journal* 42, 557-629.

Kaplow, Louis (1994), "The Value of Accuracy in Adjudication: An Economic Analysis," *Journal of Legal Studies* 23, 307-401.

Kaplow, Louis, and Steven Shavell (1994), "Accuracy in the Determination of Liability," *Journal of Law and Economics* 37, 1-15.

Kaplow, Louis, and Steven Shavell (1996), "Accuracy in the Assessment of Damages," *Journal of Law and Economics* 39, 191-210.

Keeton, Robert, Lewis Sargentich, and Gregory Keating (2004), *Tort and Accident Law*, 4th ed., St. Paul, MN: West.

Klein, David (2002), *Making Law in the United States Courts of Appeals*, Cambridge, U.K.: Cambridge University Press.

Landes, William (1971), "An Economic Analysis of the Courts," *Journal of Law and Economics* 14, 61-107.

Llewellyn, Karl (1960), *The Common Law Tradition*, Boston, MA: Little Brown.

Merryman, John (1985), *The Civil Law Tradition*, Stanford, CA: Stanford University Press.

Mullainathan, Sendhil, and Andrei Shleifer (2005), "The Market for News," *American Economic Review* 95, 1031-1053.

Pinello, Daniel (1999), "Linking Party to Judicial Ideology in American Courts: A Meta-analysis," *The Justice System Journal* 3, 219-254.

Ponzetto, Giacomo and Patricio Fernandez (2006), "Case Law vs. Statute Law: An Evolutionary Comparison," Mimeo, Harvard University.

Posner, Richard (1972), "A Theory of Negligence," *Journal of Legal Studies* 1, 29-96.

Posner, Richard (1990), *Cardozo: A Study in Reputation*, Chicago: the University of Chicago Press.

Posner, Richard (2005), "Judicial Behavior and Performance: an Economic Approach," *Florida State University Law Review* 32, 1259-1279.

Priest, George, and Benjamin Klein (1984), "The Selection of Disputes for Litigation," *Journal of Legal Studies* 13, 31-50.

Revesz, Richard (1997), "Environmental Regulation, Ideology, and the D.C. Circuit," *Virginia Law Review* 83, 1717-1772.

Shavell, Steven (1987), *Economic Analysis of Accident Law*, Cambridge, MA: Harvard University Press.

Shavell, Steven (2005), "Optimal Discretion in the Application of Rules," Mimeo.

Shavell, Steven (2006), "The Appeals Process and Adjudicator Incentives," *Journal of Legal Studies* 35, 1-29

Stone, Julius (1985), *Precedent and the Law: Dynamics of Common Law Growth*, Sydney: Butterworths.

Sunstein, Cass, David Schkade, and Lisa Michelle Ellman (2004), "Ideological Voting on Federal

Courts of Appeals: a Preliminary Investigation," *Stanford Law Review* 90, 301-354.

Viscusi, Kip (1988), "Pain and Suffering in Product Liability Cases: Systematic Compensation or Capricious Awards?," *International Review of Law and Economics* 8, 203-220.

Viscusi, Kip (1998), "The Social Costs of Punitive Damages Against Corporations in Environmental and Safety Torts," *Georgetown Law Journal* 87, 285-345.

Wyzanski, Charles (1972), "An Activist Judge: *Mea Maxima Culpa, Apologia Pro Vita Mea*," *Georgia Law Review* 7, 202-225.

Yildiz, Muhamet (2004), "Waiting to Persuade," *Quarterly Journal of Economics* 119, 223-248.