Stare Decisis and Judicial Log-Rolls: A Gains-from-Trade Model[*]

Charles M. Cameron[a], Lewis A. Kornhauser[b] , and Giri Parameswaran[c]

[a]Princeton University, ccameron@princeton.edu

[b]New York University School of Law, lewis.kornhauser@nyu.edu

[c]Haverford College, gparames@haverford.edu

Abstract

The practice of horizontal *stare decisis* requires that judges occasionally decide cases 'incorrectly'. What sustains this practice? Given a heterogeneous bench, we show that the *increasing differences in dispositions* property of preferences generates gains when judges trade dispositions over the case space. These gains are fully realized by implementing a compromise rule – *stare decisis*. Absent commitment, we provide conditions that sustain the compromise in a repeated game. When complete compromises become unsustainable, partial compromises still avail. Moreover, judges may prefer to implement partial compromises even when perfect ones are sustainable. Thus, *stare decisis* is consistent with a partially-settled,partially-contested legal doctrine.

1 Introduction

A desideratum of a well functioning judiciary is that courts decide like cases similarly. In the common law tradition, the body of existing case law (i.e. 'precedent') constitutes a source of law, and under the doctrine of *stare decisis*[1], courts should decide instant cases following the results or principles laid down in prior cases. Indeed, many areas of law are considered 'settled' and cases are decided consistently, notwithstanding the heterogeneity in judges' beliefs about how they ought to decide cases. However, in other areas of law, there is less agreement between courts, and the same case may be decided differently depending on which judge is presiding. To give just a few examples, courts continue to disagree about what constitutes an 'offer' or how to determine 'unconscionability' in contract law, how to assess 'materiality' in securities law, or what constitutes 'probable cause' in the context of search and seizure.[2] And these disagreements persist even after hundreds of cases have been decided. In these contexts, the law resembles a 'standard' in which judges exercise discretion over how cases are decided within some bounds, rather than a 'rule' which prescribes how the case should be decided.

In this paper, we take up the question of why heterogeneous judges agree to decide cases according to a consistent rule in some areas of law, but leave the law persistently unsettled in others. Our analysis begins with the recognition, especially in the context of horizontal *stare decisis*[3], that there is nothing that compels courts to follow precedent. The United States judiciary is a heterogeneous body, comprised of judges having different interpretations and

---

[1]The term *"stare decisis"* derives from the Latin phrase "Stare decisis et non quieta movere," an injunction "to stand by decisions and not to disturb the calm." The practice may vary across constitutional, statutory, and common law areas of law. *Stare decisis* rests on another practice that indicates when the instant case is governed by some (and by which) prior case. This practice also underlies the related practice of "distinguishing." For a brief discussion of *stare decisis* see Kornhauser (1998). For more extended discussions see Cross and Harris (1991), Duxbury (2008) and Levi (2013).

[2]We thank an anonymous referee for suggesting these examples.

[3]Horizontal *stare decisis* refers to the practice of following the prior decisions of a given court. Vertical *stare decisis*, by contrast, refers to the practice of following the decisions of superior courts in a hierarchical judicature. Lower courts that fail to follow the precedent established by superior courts may soon see their decisions overturned, which may in itself create an incentive to obey. In this paper, we focus on horizontal *stare decisis*, where such a hierarchy is absent.

conceptions of the law. In principle, each judge is free to dispose of cases as she sees fit. The practice of horizontal *stare decisis*, then, is voluntary; it reflects a tacit agreement, among judges to decide cases in a consistent manner, even if doing so requires them to occasionally decide cases in a manner that is contrary to their ideal.

We develop a theory that explains why judges may voluntarily apply a common compromise rule when they are free to decide cases by their own lights. We make two important contributions. First, we show that cooperation between judges is a natural consequence of a simple and plausible feature of judicial preferences, which we call the *increasing differences in dispositional values (IDID)* property. We demonstrate that such preferences create opportunities for gains-from-trade, and that these gains are maximized when judges settle on applying a common legal rule to decide cases. Second, we introduce the notion of partial *stare decisis*, which is novel in the literature. We show that commitment problems may preclude judges from fully realizing all gains-from-trade, but that partial gains may nevertheless be realized. This provides a novel account for the sustained existence of areas of law that are partially settled and partially contested.

To study the political logic of *stare decisis*, we analyze a model in which a heterogeneous court decides an infinite stream of cases that present a common issue. A case is summarized by a number on the real line that captures salient facts, e.g. the speed at which the defendant was driving. Each case is decided by a single judge selected at random from the bench.[4] Judges have preferences over case dispositions characterized by a threshold that separates cases that would ideally be decided for the plaintiff from those that would ideally be decided for the defendant. Importantly, we assume that these preferences satisfy the *increasing differences in dispositional value (IDID)* property, which states that the utility differential from deciding

---

[4]We thus offer a stylized model of intermediate courts of appeal in which a panel of judges is drawn from a wider bench. If the judges on the court divided into two factions, each panel (with an odd number of judges) would, absent panel effects, decide as the faction of the majority of its members would. The practice of drawing a panel from a bench is common on intermediate courts of appeal in the United States as well as on appellate courts in other legal systems. Many supreme courts outside the United States and many international courts have a similar practice.

cases 'correctly' rather than not, increases the further the case is from the threshold; it is more costly to incorrectly decide 'clear-cut' cases than 'contestable' ones. We introduce heterogeneity into the model by assuming that judges are drawn from one of two factions ('L' and 'R'), which are distinguished by the location of the threshold case. Although judges may only decide the cases before them, they receive utility from the disposition of every case, regardless of which judge decided it.

These simple and reasonable assumptions on judicial preferences are sufficient to provide incentives for judges to adhere to compromise rules. Given preference heterogeneity, there will be a set of cases (the 'conflict zone') in which, no matter how the case is decided, some judge(s) will be disappointed. By the *IDID property*, the relative cost of an adverse decision varies over the set; the cost to L faction judges of having cases close to the L-threshold decided incorrectly is small relative to the cost of incorrectly deciding cases close to the R-threshold (far from the L-threshold). The opposite is true for R faction judges. This creates an opportunity for gains-from-trade. Although each judge *could* dispose of cases according to her ideal, there are mutual gains from judges instead implicitly agreeing to dispose of cases in a way that assigns the losses to the least-cost abater. When all such gains are exhausted, the disposition of cases is shown to be independent of the identity of the presiding judge – cases are decided as if all judges adhered to the same consistent rule. In section 1.1, we present a simple example to illustrate this gains-from-trade logic.

Implicit in the logic of trade was that both judges (simultaneously) decided cases in a way that was mutually beneficial. But, in our framework, judges are assumed to decide cases sequentially, one at a time. If side payments were permitted, this would not matter, since the non-presiding judge could always make a transfer to the presiding judge that incentivized the Pareto optimal outcome. However, in this model, as in the real world, such payments are not available. We show that, under certain conditions, trade can nevertheless be sustained provided that judges are incentivized to behave reciprocally. When presiding, L faction

3

judges choose to 'incorrectly' decide cases close to their threshold in the expectation that R faction judges will reciprocate by 'incorrectly' deciding cases close to the R threshold (far from the L threshold) when they are presiding over future cases. But judges have no formal commitment device. Cooperation is sustainable – i.e. the promise to reciprocate is credible – only if the immediate losses stemming from the compromise are not too large for either faction, and that judges are sufficiently patient that the future gains outweigh current losses.

Even if the lack of commitment prevents judges from fully realizing the gains from trade, they may nevertheless be able to commit to realizing partial gains. Start from a baseline of no compromise ('Autarky') and consider the effect of each faction making incremental compromises towards the middle. By the *IDID* property, the early increments are least costly to the compromising faction, and provide the largest gains to the other faction. Latter increments achieve (relatively) smaller gains and at higher cost. Under partial *stare decisis*, the judges agree to a compromise over a subset of cases – those for which the benefits of cooperation are large and the costs small – but continue to dispose of the remaining cases according to their ideal. Since the costs of adhering to such a compromise are small, so is the temptation to defect, making an equilibrium with partial compromises easier to sustain. Indeed, since the largest gains come from the initial compromises, judges may be able to realize the lion's share of potential gains-from-trade under partial *stare decisis*, even if they cannot realize all gains.

A partial *stare decisis* equilibrium illuminates a pervasive and important feature of common law adjudication. In many instances, the law governing some behavior persistently remains unsettled over some range of conduct. One might understand the law of negligence in this way; the law sets bounds on behavior that, depending on the view of a jury may be either negligent or not but no reasonable jury could conclude on the one hand that certain extreme behavior is non-negligent or other behavior, extreme in the other direction, is negligent. Our model offers two complementary reasons why this legal structure emerges

and persists. On the one hand, we show that for a range of parameters, judges cannot credibly commit to implementing a single rule (i.e. complete *stare decisis*), but may still find it beneficial to narrowing the scope of unsettled law (i.e. partial *stare decisis*), since this insures them against adverse outcomes in the most costly cases. On the other hand, we show (in Proposition 3) that either faction may prefer to implement a partial *stare decisis* equilibrium even when a full stare decisis equilibrium is sustainable, if in doing so, the expected cut-point implemented is more favorable.

Our model makes several predictions about the nature of legal rules. First, we show that judges are more likely to apply consistent rules in areas of law that are highly salient or which are frequently litigated. By contrast, when these conditions are not met, the law is more likely to be characterized by partial *stare decisis*. Second, we show, somewhat counter-intuitively, that judges are more likely to apply a consistent rule when the bench is highly polarized, than when it is ideologically cohesive. Third, we show that the location of compromise rules is monotone in the relative strength of the factions, such that the expected threshold policy moves further to the left as the left faction of the bench becomes more dominant.

Important in the above discussion was the distinction between dispositions and rules. A case *disposition* is a determination of which party to a dispute prevails. A *rule* is a menu that specifies a disposition for every potential case that may arise. It contemplates how a judge would decide a variety of counter-factual cases. When deciding cases, judges choose dispositions. They decidedly do not choose rules — since a rule determines the disposition for not only the instant case, but all future cases as well. (As we have already argued, judges are only empowered to determine the case before them. They cannot commit future courts to decide cases one way or other.) Of course, each individual judge may dispose of cases in a manner consistent with a particular rule. This is wholly different from the judge choosing a rule to be applied to all future cases by all future judges. Indeed, it is precisely the goal of this paper to ask when heterogeneous judges will each individually determine to apply the

same rule, so that it appears that a consistent rule is being applied by the court as a whole. Throughout this paper, we will refer to judges as 'applying' rules rather than 'choosing' rules to emphasize the distinction between the court choosing rules and disposing of cases in a manner consistent with a rule.

An important element of the model is it's grounding on cases and case dispositions. In the model, judges decide concrete cases rather than announce abstract policies. Of course, deciding cases is what judges actually do. But the case-based nature of the model is important for more than verisimilitude. Our analysis investigates the sustainability of *stare decisis* rather than simply assume it. But adherence to *stare decisis* means deciding specific cases one way rather than another. Similarly, defection from *stare decisis* means deciding specific cases in an inappropriate way. Moreover, punishing deviations requires deciding future cases in a way that deviators find punitive. Hence, commitment and the enforcement of commitment in a model of *stare decisis* are inextricably tied to cases and case dispositions. Finally, partial *stare decisis* – following *stare decisis* for one subset of cases but not another – cannot even be conceptualized without reference to cases. Fortunately, the apparatus for modeling cases and case dispositions is relatively straightforward, as explained below.

The paper is organized as follows. In the remainder of this section, we provide a simple motivating example that illustrates how *stare decisis* allows an ideologically diverse bench to realize gains from trade. We also briefly discuss relevant literatures. Section 2 lays out the model, including a formal representation of cases and of judicial utility. Section 3 examines the stage game, establishing that autarky is the unique Nash equilibrium in a one-shot setting. However, autarky is inefficient. Thus the judges find themselves in a continuous action prisoner's dilemma. Section 4 investigates the conditions under which a practice of *stare decisis* is sustainable absent commitment but with an infinite stream of cases. Section 5 considers partial *stare decisis* regimes, and investigates when they are preferable to complete *stare decisis* regimes.

6

1.1 Judicial Logrolling: An Illustration

In this subsection, we illustrate the logic of gains-from-trade through the following example. Let x denote the intensity of some behavior, which we understand to be the facts of a case. There are two judges, L and R, each of whose preferences over case dispositions is characterized by the location of a threshold case. Let the thresholds be 0 and 1 for judge L and R, respectively. We interpret this as saying, given a case x, that L would ideally find for the defendant whenever $x < 0$ and for the plaintiff otherwise. By contrast, R would ideally find for the defendant whenever $x < 1$ and for the plaintiff otherwise. The judges agree on the ideal disposition when $x < 0$ or $x > 1$, but disagree over the ideal disposition when $x \in [0, 1]$.

Suppose judicial preferences satisfy:

$$u^i(x) = \begin{cases} 0 & x \text{ decided 'correctly'} \\ -|x - \overline{y}_i| & x \text{ decided 'incorrectly'} \end{cases}$$

where $\overline{y}_i \in \{0, 1\}$ is judge i's threshold. In words, a judge receives utility 0 whenever the case disposition accords with her ideal. If the case is disposed 'incorrectly' she suffers a loss that is increasing in the distance of the case from the threshold. Hence, it is costlier for cases farther from a judge's threshold to be decided incorrectly, than cases nearer to the threshold; preferences satisfy the *IDID* property. The linearity of the loss function is purely for simplicity, and in the main section we generalize to any preferences satisfying the *IDID* property.

Suppose there are two equally likely potential cases located at $\frac{1}{4}$ and $\frac{3}{4}$. Both cases are in the region of conflict, so no matter how a case is decided, one of the judges will be disappointed. Suppose each judge is equally likely to be selected to decide the case. Hence, there are four equally likely case-judge pairs: $\left(\frac{1}{4}, L\right)$, $\left(\frac{1}{4}, R\right)$, $\left(\frac{3}{4}, L\right)$ and $\left(\frac{3}{4}, R\right)$.

Table 1: Stage Payoffs Under Autarky and *Stare Decisis*.

Autarky					Stare Decisis				
(Case, Judge)	$\left(\frac{1}{4},L\right)$	$\left(\frac{1}{4},R\right)$	$\left(\frac{3}{4},L\right)$	$\left(\frac{3}{4},R\right)$	(Case, Judge)	$\left(\frac{1}{4},L\right)$	$\left(\frac{1}{4},R\right)$	$\left(\frac{3}{4},L\right)$	$\left(\frac{3}{4},R\right)$
Disposition	P	D	P	D	Disposition	D	D	P	P
L-utility	0	$-\frac{1}{4}$	0	$-\frac{3}{4}$	L-utility	$-\frac{1}{4}$	$-\frac{1}{4}$	0	0
R-utility	$-\frac{3}{4}$	0	$-\frac{1}{4}$	0	R-utility	0	0	$-\frac{1}{4}$	$-\frac{1}{4}$

The dispositions P and D indicate outcomes favoring the plaintiff and defendant, respectively.

First, consider outcomes under 'autarky', where judges dispose of cases according to their ideal legal rule. If so, L will decide both cases for the plaintiff and R will decide both cases for the defendant. With probability $\frac{1}{2}$, each judge gets her ideal outcome, and with probability $\frac{1}{2}$ she faces an adverse outcome. Importantly, the costs of these adverse outcomes vary. If R decides the case, then L's payoff is $-\frac{1}{4}$ if the case is $x = \frac{1}{4}$, whereas his payoff is $-\frac{3}{4}$ if $x = \frac{3}{4}$. Similarly, if L decides the case, then R's payoff is $-\frac{3}{4}$ if $x = \frac{1}{4}$ and $-\frac{1}{4}$ if $x = \frac{3}{4}$. It is costlier for each judge to receive an adverse outcome when the case is farther from their threshold than when it is nearer. Each judge's expected payoff is $-\frac{1}{4}$. The payoffs to each judge are represented in left panel of Table 1.

Now, suppose the judges affected a compromise policy, where each finds for the defendant when $x < \frac{1}{2}$ and each finds for the plaintiff when $x > \frac{1}{2}$. The compromise requires L to decide cases in the region $\left(0, \frac{1}{2}\right)$ 'incorrectly' by finding for the defendant when he would ideally find for the plaintiff. Similarly, it requires R to decide cases in the region $\left(\frac{1}{2}, 1\right)$ 'incorrectly' by finding for the plaintiff when she would ideally find for the defendant. Under this compromise regime, the outcomes depend solely on the facts of the case, and are independent of the identity of the presiding judge. L gets her desired outcome whenever the case is $x = \frac{3}{4}$, and the adverse outcome whenever the case is $x = \frac{1}{4}$. The opposite is true for R. Each judge's expected payoff is $-\frac{1}{8}$, which is an improvement over autarky. As with autarky, each judge gets their preferred outcome with probability $\frac{1}{2}$. However, now, adverse outcomes are assigned to the judge who suffers the lowest disutility.

The outcome under *stare decisis* represents a Pareto improvement over autarky. It is enabled

by the judges 'trading' case dispositions. In both scenarios, L finds for the plaintiff when $x = \frac{3}{4}$ and R finds for the defendant when $x = \frac{1}{4}$. However, under *stare decisis*, L agrees to find for the defendant when $x = \frac{1}{4}$, even though he would ideally find for the plaintiff, and R agrees to find for the plaintiff when $x = \frac{3}{4}$, even though she would ideally find for the defendant. L's concession to R involves a utility cost to L that is outweighed by the utility gain from R's concession to L. After the trade, it is as if cases are decided by the judge who would face the largest relative cost from an adverse outcome, regardless of which judge actually hears the case.

To see the role played by the *IDID* property, consider the following alternative example. Suppose a correct disposition yields utility of 0 but an incorrect disposition yields utility of -1 regardless of the case location. Using these payoffs, it can be seen in the example that each judge's expected payoff is the same (and equal to $-\frac{1}{2}$) under both scenarios. There are no longer gains from trade. Judges only care about whether cases are decided 'correctly' or not. They do not value 'correct' versus 'incorrect' dispositions differently across the case-space. *Stare decisis* no longer allows the judges to trade small-loss dispositions in order to gain high-value ones. As a result, the *stare decisis* regime no longer offers Pareto improvements over autarky. This example underscores the importance of the IDID property in this model of stare decisis.

In this example, we have not constructed an equilibrium; we have simply demonstrated how the stage game payoffs work, and the potential for gains-from-trade. We take up the question of the sustainability of log-rolling in detail, below.

1.2 Related Literature

A vast literature on stare decisis exists but almost none of it addresses the practice's sustainability.[5] Two papers, however, consider mechanisms related to that analyzed here.

O'Hara (1993) offers an informal argument that includes many elements of our formal model. She considers a pair of ideologically motivated judges who engage in what she calls "non-productive competition" over the resolution of a set of cases that should be governed by a single rule.[6] The judges hear cases alternately. She recognizes that they each might do better by adhering to a compromise rule. She characterizes their interaction as a prisoner's dilemma and then relies on the folk theorem to argue for the existence of a mutually preferred equilibrium. Because her argument is informal, the nature of the exchange between the two judges is unclear. Similarly, she cannot provide comparative statics for the degree of polarization or the relative sizes of the factions (i.e., the relative probability that each judge will hear a case).[7] Nor does she consider partial *stare decisis*.

Rasmusen (1994) too shows that repeated play can sustain *stare decisis*. But his model has a very different structure. He considers an infinite sequence of judges, each of whom hears $n + 1$ cases, only one of which is a case of first impression. In his model, *stare decisis* is an agreement by each judge to follow the precedents of prior judges in the n non-first impression

---

[5]Much of the literature there is normative. Relevant literature includes Goodhart (1930), Levi (2013), Llewellyn (2012), Cross and Harris (1991), MacCormick and Summers (1997) and Duxbury (2008), as well as Postema (1986) and Schauer (1987) on the desirability of the practice. A sizable literature in Political Science searches for empirical evidence on the extent of stare decisis on the U.S. Supreme Court. See Spaeth and Segal (2001), Knight and Epstein (1996), Brisbin (1996), Brenner and Stier (1996), and Songer and Lindquist (1996). Several authors investigate reasons why a court might adopt a practice of *stare decisis*. See for example Kornhauser (1989), Bueno de Mesquita and Stephenson (2002), Blume and Rubinfeld (1982), Heiner (1986) and Gely (1998).

[6]She notes that, in other instances, the judges might specialize, each generating law in a different area of law and following the precedent of the other judge on other areas of law.

[7]Several formal models of courts simply assume a practice of *stare decisis*; these models treat horizontal *stare decisis* as an exogenous constraint on judges' behavior (Jovanovic (1988), Kornhauser (1992)). In other words, these models assume a commitment device whereby one generation of judges may bind the hands of its successors. Jovanovic presents a very abstract model in which he shows that ideologically diverse judges would do better in terms of both *ex ante* and *ex post* efficiency if they were constrained by a rule of *stare decisis*. Our model identifies conditions under which rational, self-interested judges might in fact successfully adopt such a rule.

cases, with the expectation that future judges will follow his precedent in the $(n+1)^{th}$ case. By contrast, under autarky, the judge ignores all existing precedents and decides all $n+1$ cases by his own lights. The practice of *stare decisis* follows from the assumption that judges gain more utility from having their ideal opinion authored by other judges than authoring the same opinion themselves. Then, under *stare decisis*, each judge will have the same number of cases decided according to his ideal as under autarky, but will get the added benefit of having most of those opinions written by other judges.

Rasmussen's model is similar to ours in that the practice of *stare decisis* is sustained as an equilibrium of a repeated game. However, the mechanisms that drive the results differ in several important ways. First, the result in his model relies crucially on the non-consequentialist, expressive preferences of judges (who care about the identity of the opinion writer in addition to the content of the opinion). As Rasmussen recognizes (on p.74), with standard consequentialist preferences (which, in his model, amount to setting $y = x$), 'complete judicial breakdown' (i.e. autarky) is the unique equilibrium of the game. By contrast, judicial preferences in our model are consequentialist. Judges simply care about how cases are disposed, independent of other factors. The gains from trade in our model stem from the feature of preferences that judges feel more strongly the outcomes of certain cases over others. By contrast, in Rasmussen's model, the gains come entirely from judges' preferences over the identity of the opinion writer. Second, the logic of gains-from-trade does not play a particular role in Rasmussen's model. *Stare decisis* amounts to dividing the law into distinct issue-spaces, one under the control of each judge. The assignment of issue-spaces to judges is random — there is no sense in which judges trade issues according to their preferences. As such, it should be clear that the mechanism that drives compromise in Rasmussen's model is conceptually quite distinct from ours.

There are other ways in which Rasmussen's model differs from ours. His is an overlapping generations model of judges, each having distinct preferences, whereas ours is an infinite

horizon model consisting of two factions of judges whose preferences remain constant over time. Rasmussen's model, because of the structure of preferences, does not admit a measure of ideological polarization, and as such, cannot consider the comparative statics that we explore. Finally, Rasmussen does not allow for the possibility of partial *stare decisis*.

2   The Model

There are two factions of judges, L and R, that form a bench. In each period $t = 1, 2, ...$, a single judge hears a case $x_t \in X \subset \mathbb{R}$. A *case x* connotes an event that has occurred, for example, the level of care exercised by a manufacturer. Cases are drawn from a common knowledge distribution $F(x)$. We assume F is continuous, has bounded support, and admits a density $f(x)$. When a case arrives before the court, a judge is chosen at random to decide it. The chosen judge is from the L faction with probability p, and from the R faction with probability $1-p$.[8] The variable p indicates the probability that the L faction holds power, but it can also be seen as a measure of the *heterogeneity* of the bench. More precisely the bench is most heterogeneous when $p = \frac{1}{2}$ and becomes less heterogeneous (more homogeneous) as p approaches 0 or 1.

2.1   Dispositions, Rules, and Cut-points

A judicial *disposition $d \in D = \{0, 1\}$* of a case determines which party prevails in the dispute between the litigants. Judges dispose of cases by applying a legal rule. A legal rule r maps the set of possible cases into dispositions, $r : X \to D$. Let $\mathcal{X} = 2^X$ be the space of possible rules. We focus on an important class of legal rules, *cut-point-based* doctrines, which take

---

[8]Suppose the bench is composed of n judges, n^L of whom belong to the L faction and n^R of whom belong to the R faction. Then one can view $p = \frac{n^L}{n}$ and $1 - p = \frac{n^R}{n}$.

the form:

$$r\left(x;y\right) = \begin{cases} 1 & \text{if } x \geq y \\ 0 & \text{otherwise} \end{cases}$$

where y denotes the cut-point. For example, in the context of negligence, the defendant is not liable if she exercised at least as much care as the cut-point y.[9] If $X = [\underline{x}, \overline{x}]$, then the space of cut-point rules is $\mathcal{X}^C = \{\{[\underline{x}, y), [y, \overline{x}]\} \mid y \in X\}$. It should be clear that rules and cases live in entirely different spaces. The special structure of cut-point rules allows us to summarize them in terms of a threshold in case-space.

Judges from different factions differ in their assessment of the ideal rule. Formally, the ideal cut-point for a judge from faction $j \in \{L, R\}$ is $\bar{y}^j$, where $\bar{y}^L < \bar{y}^R$. The logical structure of cut-point rules implies both consensus and conflict between judges from the two factions. For a case $x < \bar{y}^L$, both judges agree that the appropriate resolution of the case is '0'. Similarly, when $x > y^R$, the two judges agree that the appropriate resolution of the case is '1'. Only when $x \in [\bar{y}^L, \bar{y}^R]$ do the judges disagree: $y^L < x < y^R$ implies that the L judge believes the appropriate resolution of x is '1' while the R judge believes the appropriate resolution is '0'. We are particularly interested in this *conflict region.*

To simplify the analysis, we assume $[\bar{y}^L, \bar{y}^R] \subset supp\left(F\right)$, which implies that F is strictly increasing over the region of conflict. In footnote 17 we show that this assumption is benign in that it merely rules out trivial multiplicities of equilibria. Finally, we parameterize the degree of *polarization* ρ by the fraction of cases in the region of conflict; $\rho = F\left(\bar{y}^R\right) - F\left(\bar{y}^L\right)$.

---

[9]Other examples include allowable state restrictions on the provision of abortion services by medical set providers; state due process requirements for death sentences in capital crimes; the degree of procedural irregularities allowable during elections; the required degree of compactness in state electoral districts; and the allowable degree of intrusiveness of police searches. Many other examples of cut-point rules may suggest themselves to the reader.

2.2 Dispositional Utility

The governing judge's disposition of the case affords utility to *all* judges, reflecting their conceptions of the 'correct' disposition of the case. Stage utility for judge i in each period t is determined by the dispositional utility function[10]:

$$
u(d; x, \overline{y}^i) = \begin{cases} g(x; \overline{y}^i) & \text{if the disposition is 'correct'} \\ l(x; \overline{y}^i) & \text{if the disposition is 'incorrect'} \end{cases}
$$

where $\overline{y}^i$ connotes judge i's most-preferred standard. $g(x; \overline{y}^i)$ denotes judge i's dispositional utility over gains – when the judge in power disposes of the case 'correctly'; that is, if the ruling judge reaches the same disposition that judge i, applying her most-preferred cut-point $\overline{y}^i$, would have reached. Conversely, $l(x; \overline{y}^i)$ denotes judge i's dispositional utility if the judge in power disposes of the case 'incorrectly'. Both the ruling judge and the non-ruling judges receive utility each period.

Since the judges' choices are binary, what is salient to the judges is the utility differential between having a case decided correctly versus incorrectly. For notational convenience, we denote by $\eta(x; \overline{y}^i) = g(x; \overline{y}^i) - l(x; \overline{y}^i)$, the net benefit of correctly disposing of case x. We assume that net dispositional utility satisfies two properties:

1. $\eta(x; \overline{y}^i) \geq 0$ for all x (i.e. $g(x; \overline{y}^i) \geq l(x; \overline{y}^i)$). It is better that a case be correctly disposed rather than not;

2. $\eta(x; \overline{y}^i)$ is strictly decreasing for $x < \overline{y}^i$, strictly increasing for $x > \overline{y}^i$, and achieves a minimum at $x = \overline{y}^i$. The net benefit of correctly deciding a case increases the further the case is from the threshold.

The first property is straightforward. The second property is a formal statement of the *IDID*

---

[10] Formally, judge i views a disposition d of case x as 'correct' if $d = r\left(x; \overline{y}^i\right)$, and 'incorrect' otherwise.

property.[11] The *IDID* property amounts to asserting that $-\eta\left(x;\overline{y}^i\right)$ has a single peak at $\overline{y}^i$. We assume that the functions g and η are continuous, which ensures they are integrable with respect to the distribution function F. Beyond these, we make no further assumptions about the properties of η; in particular, η is not assumed to be convex — judges need not be 'risk averse' over case dispositions.

2.3   Strategies and Equilibrium

A history of the game h_t indicates all the prior cases and the dispositions afforded them by the ruling judges. The set of possible histories at the beginning of time t is thus $H_t = X^{t-1} \times D^{t-1}$. A *behavioral strategy* σ^j for ruling judge j is a mapping from the set of possible histories and the set of possible current cases into the set of dispositions $\sigma^j : H_t \times X \to D$; a strategy σ^j for judge j is a history dependent selection of a legal rule. We focus on symmetric strategies, so that all judges from the same faction choose the same strategy.

Let $\left(\sigma^L, \sigma^R\right)$ be a pair of strategies for L and R-type judges, respectively. Let $V^i\left(h_t\right)$ denote the expected lifetime utility of judge i after history h_t, given strategies $\left(\sigma^L, \sigma^R\right)$. Lifetime utility satisfies the Bellman equation:

$$V^i\left(h_t\right) = p \int \left[u\left(\sigma^L\left(h_t, x\right); x, \overline{y}^i\right) + \delta V^i\left(h_t, x, \sigma^L\left(h_t, x\right)\right)\right] dF\left(x\right) +$$

$$+ \left(1 - p\right) \int \left[u\left(\sigma^R\left(h_t, x\right); x, \overline{y}^i\right) + \delta V^i\left(h_t, x, \sigma^R\left(h_t, x\right)\right)\right] dF\left(x\right) \qquad (1)$$

where $\delta \in [0, 1)$ is the common discount factor. The lifetime utility V^i is simply the expected discounted stream of stage utilities that judge i receives, given the strategies chosen by each

---

[11]Although not strictly necessary, the reader might find it helpful to add the following additional structure. Suppose g is weakly increasing in the distance from the threshold. This implies that judges receive higher utility from correctly deciding cases that are further from the threshold. Similarly, suppose l is weakly decreasing in the distance from the threshold. Judges receive larger disutility from incorrectly deciding cases that are further from the threshold. If the threshold represents the case where the judge is indifferent, then as the distance from the threshold increases, the ideal disposition becomes more 'clear cut', and the utilities respond accordingly.

faction.

A pair of strategies $\left(\sigma^L, \sigma^R\right)$ is a sub-game perfect equilibrium if, for every $x \in X$ and after every history $h_t \in H_t$:

$$\sigma^i\left(h_t, x\right) \in \arg\max_{d \in \{0,1\}} \left\{ u^i\left(d; x, \bar{y}^i\right) + \delta V^i\left(h_t, x, d\right) \right\} \tag{2}$$

We solve for symmetric sub-game perfect equilibria.

At this juncture, we again clarify the nature of the choices that the judges make. In each period, one judge will choose the disposition in the case before her. This is the only actual choice that is made. Since a strategy specifies how a given judge should dispose of every possible case, specifying a strategy for a judge amounts to specifying a rule that the judge applies to decide the case before him. In principle, the judge may apply different rules after different histories. Importantly, the judge does not choose a rule for the court. Indeed, a rule for the court is never chosen. However, if all judges' equilibrium strategies select the same rule to be applied, then the individual decisions of judges in disposing of cases will be consistent with the operation of a common rule.

2.4   The Practice of *Stare Decisis*

The norm of *stare decisis* is in effect if all judges would dispose of the same case in the same way. For example, suppose at time t, the strategies direct judges from factions L and R to adjudicate cases according to cut-points z_t^L and z_t^R (respectively), where $z_t^L < z_t^R$. Then, all judges will identically dispose of cases $x < z_t^L$ and $x > z_t^R$. The law in those regions is *settled*; there is common agreement about how such cases will be decided. By contrast, different judges will differently dispose of cases in the region $\left[z_t^L, z_t^R\right]$. In this region, the law remains *contested.* There is an established norm of *stare decisis* to the extent that cases arise in the settled region of the case space.

Two special cases are worth noting. First, if $z_t^L = \overline{y}^L$ and $z_t^R = \overline{y}^R$, then the period t regime is characterized by 'autarky'; each judge simply applies their ideal legal rule, and there is no attempt at implementing a consistent, compromise legal rule.[12] The contested region of law comprises the entire zone of conflict between the two factions. Second, if $z_t^L = z_t^R$, then the period t regime is characterized by complete *stare decisis*; the judges would decide all cases in the same way. The law is settled over the entire case space.

As we show in the following section, it is natural to expect $\overline{y}^L \leq z_t^L \leq z_t^R \leq \overline{y}^R$. If so, the cut-points $\left(z_t^L, z_t^R\right)$ divide the region of conflict $\left[\overline{y}^L, \overline{y}^R\right]$ into three sub-regions. In region $\left[\overline{y}^L, z_t^L\right]$, L faction judges are required to 'compromise' by returning the disposition favored by R faction judges. Similarly, in region $\left[z_t^R, \overline{y}^R\right]$, R faction judges are required to compromise by returning the disposition favored by L faction judges. The norm of *stare decisis* is maintained in these regions. By contrast, in the region $\left(z_t^L, z_t^R\right)$, the law is contested. Judges of neither type compromise, and instead each disposes of cases according to their preferred rule. The strategy pair $\left(z^L, z^R\right)$ suggests the operation of a standard, where judges are able to exercise discretion when deciding cases, within some bounds.

In an actual practice of (horizontal) *stare decisis*, judges adhere to the decisions rendered by prior courts. The legal literature disagrees about what in the prior decision "binds" the judge: the disposition, the announced rule, or the articulated reasons.[13] Our formalism seems consistent with judges respecting either a previously announced rule z^C or respecting the dispositions of prior decisions that give rise to a coherent rule as in the model in Baker and Mezzetti (2012).

---

[12]Technically, the norm of *stare decisis* still prevails for all cases arising outside the region of conflict. However, cooperation over this region is trivial, given that all judges agree about the ideal disposition of cases in these regions. In this paper, we focus attention on the ability of judges to sustain norms of cooperation in cases where their ideal dispositions diverge.

[13]For some discussion see Kornhauser (1998).

3 The Stage Game

3.1 Payoff functions in the Stage Game

In this section we analyze the stage game. As we have previously noted, we take as primitive the judges' preferences over dispositions $u\left(d; x, \bar{y}^i\right)$. Note that we can write:

$$u\left(d; x, \bar{y}^i\right) = g\left(x; \bar{y}^i\right) - \mathbf{1}\left[d \neq r\left(x; \bar{y}^i\right)\right] \eta\left(x; \bar{y}^i\right)$$

Since judges dispose of cases by applying rules, we can evaluate different rules by the expected utility that the judge derives, given the dispositions that are induced by the rule and the distribution over cases. Let $v^i\left(z\right) = \int_{x \in X} u\left(r\left(x; z\right); x, \bar{y}^i\right) dF\left(x\right)$ denote judge i's *policy utility* from having cut-point rule z implemented. We have[14]:

$$v^i\left(z\right) = E\left[g\left(x; \bar{y}^i\right)\right] - \left|\int_{\bar{y}^i}^{z} \eta\left(x; \bar{y}^i\right) dF\left(x\right)\right| \tag{3}$$

. Additionally, since judges from the different factions may apply different rules, the judge's utility will not simply reflect the application of a particular rule z, but a pair of (potentially different) rules (z^L, z^R). We can thus think of the structure of the adjudication game as a two-stage lottery. In the first stage, the presiding judge (and hence rule) is selected, and in the second stage the case arrives and is decided. The dispositional utility function $u\left(d; x, y^j\right)$ represents judge j's *ex post* utility after both lotteries have resolved — when the case to be decided and the rule being applied are known. The policy utility function $v^i\left(z\right)$ represents the judge's *interim* utility, after the resolution of the first stage lottery — when the rule is known, but not the case to which it will be applied. Finally, the judge's *ex ante* utility $v^i\left(z^L, z^R\right)$ represents the judge's expected policy utility at the start of the game, before

---

[14]The absolute value corrects the sign of the integral if it is negatively oriented; i.e. if $\bar{y}^i > z$.

either the case or the presiding judge have been selected. We have:

$$v^i\left(z^L, z^R\right) = pv^i\left(z^L\right) + (1-p)\, v^i\left(z^R\right)$$

$$= E\left[g\left(x; \bar{y}^i\right)\right] - p\left|\int_{\bar{y}^i}^{z^L} \eta\left(x; \bar{y}^i\right) dF\left(x\right)\right| - (1-p)\left|\int_{\bar{y}^i}^{z^R} \eta\left(x; y^i\right) dF\left(x\right)\right| \quad (4)$$

which is the sum of three terms.[15] The first term is independent of the chosen legal rule, and establishes the judge's baseline utility if every case were decided in accordance with her ideal rule. The second term is the expected loss incurred when a judge from faction L decides a case differently from judge i's ideal. Suppose $\bar{y}^L < z^L < \bar{y}^R$. If judge i is herself from faction L, then this will occur when $x \in \left[\bar{y}^L, z^L\right]$; when the actual disposition is '0' although her ideal disposition would be '1'. If judge i is from faction R, then this will occur when $x \in \left[z^L, \bar{y}^R\right]$; when the actual disposition is '1' although her ideal disposition would be '0'. Similarly, the third term is the expected loss incurred when a judge from faction R decides a case differently from judge i's ideal.

Example 1. Suppose $g\left(x; \bar{y}^i\right) = 0$ for all x, and $l\left(x; \bar{y}^i\right) = -\left|x - y^i\right|$, which implies $\eta\left(x; \bar{y}^i\right) = \left|x - \bar{y}^i\right|$.[16] Further suppose $X \sim U\left[\underline{x}, \bar{x}\right]$, where $\underline{x} < \bar{y}^L < \bar{y}^R < \bar{x}$, which ensures that every case in the conflict region arises with positive probability. Denote $\varepsilon = \frac{1}{2}\left(\bar{x} - \underline{x}\right)$, the distance between the mean and boundary cases. Then (3) yields: $v^i\left(z\right) = -\frac{1}{4\varepsilon}\left(z - \bar{y}^i\right)^2$ and (4) yields: $v^i\left(z^L, z^R\right) = -\frac{1}{4\varepsilon}\left[p\left(z^L - \bar{y}^i\right)^2 + (1-p)\left(z^R - \bar{y}^i\right)^2\right]$.

3.2 Equilibria and Optima

Consider a one-shot version of the game in which each judge chooses a rule to maximize her *ex-ante* utility, taking as given the strategy of the other judge.

---

[15]Without confusion, we use $v^i\left(z\right)$ to denote judge i's policy utility if the rule z is applied, and $v^i\left(z^L, z^R\right)$ to denote the expected policy utility from a first stage lottery generating rule z^L with probability p, and rule z^R with probability $1 - p$.

[16]This very tractable utility function has been deployed by others (e.g., see Fischman (2011)). Furthermore, in other work, two of us present other arguments that recommend the choice of $g\left(x; \bar{y}^i\right) = 0$ for all x, more generally (see Cameron and Kornhauser (2017)).

Lemma 1. *In the one shot game, it is a strictly dominant strategy for each judge to apply her ideal rule.*

Lemma 1 makes the straightforward point that, absent any other incentives, the best that any judge can do is implement her own ideal policy whenever she is chosen to adjudicate a dispute. Autarky is a Nash equilibrium of the one-shot adjudication game. However, as the next Lemma shows, autarky is not Pareto optimal. The judges could do better by cooperating.

Lemma 2. *The set of complete stare decisis regimes defines the Pareto frontier. Formally, a pair of rules $\left(z^L, z^R\right)$ is Pareto optimal iff $\bar{y}^L \leq z^L = z^R \leq \bar{y}^R$.*

Lemma 2 highlights the main insight of this paper; that judges can enjoy gains from trade when judges from both factions resolve cases in the same way. This 'cooperation' amounts to the application of the same rule to all cases that arise in the conflict zone, independent of the identity of the presiding judge.[17] To build intuition for the result in Lemma 2, consider the following 'Edgeworth Box'-type diagram, which represents the preferences of both types of judges in policy-space. The square box represents all pairs of legal rules whose cutpoints lie in the region of conflict. The top-left corner of the box corresponds to Autarky, since $\left(z^L, z^R\right) = \left(\bar{y}^L, \bar{y}^R\right)$. The top-right corner of the box corresponds to an R faction judge's ideal legal rule, whilst the bottom-left corner of the box corresponds to an L faction judge's ideal rule. The 45° line is the locus of all complete *stare decisis* regimes since $z^L = z^R$. Intuitively, we should expect rules in the upper triangle, since in this region, $z^L \leq z^R$.

---

[17]The assumption that F is strictly increasing ruled out trivial multiplicities of optima. Relaxing this assumption, Lemma 2 generalizes to the claim: 'A rule pair $\left(z^L, z^R\right)$ is Pareto optimal iff $F\left(\bar{y}^L\right) \leq F\left(z^L\right) = F\left(z^R\right) \leq F\left(\bar{y}^R\right)$.' To see this, suppose F is weakly increasing and $F\left(z^L\right) = F\left(z^R\right)$ with $z^L \neq z^R$. Then the partial *stare decisis* regime $\left(z^L, z^R\right)$ is Pareto optimal because it induces the same pattern of case dispositions as the complete *stare decisis* regime $\left(z^C, z^C\right)$, where $z^L \leq z^C \leq z^R$. The strategies are observationally equivalent. Although the rule-pairs are distinct in principle, they disagree over a measure-zero set of cases, and so their distinction is trivial. Hence, restricting attention to strictly increasing F is without loss of generality. A similar insight applies to Lemma 1 which generalizes to the claim: 'If $F\left(z^i\right) = F\left(\bar{y}^i\right)$, then implementing rule z^i is a dominant strategy for judge i.'

Figure 1: An Edgeworth Box Representation of Gains From Trade
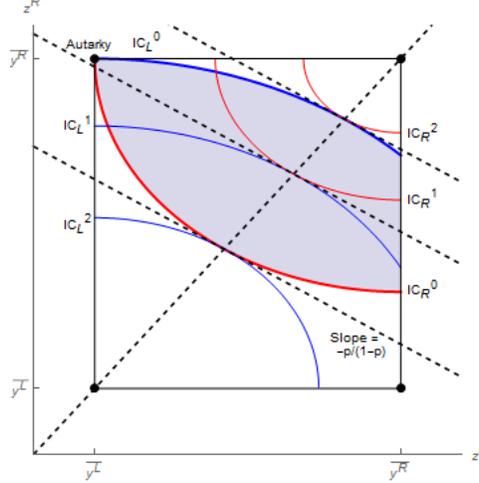
The diagram assumes F is uniform over the region of conflict.

(Rule-pairs in the lower triangle have the property that each type of judge prefers the rule implemented by the other type of judge, to their own.)

The downward sloping dashed lines have slope $\frac{-p}{1-p}$ and connect all rule-pairs that share the same expected cut-point. Indifference curves for R faction judges improve in the north-east direction, whereas indifference curves for L faction judges improve in the south-west direction. The slopes of indifference curves are given by:[18]:

$$\frac{\partial z^R}{\partial z^L}\big|_{v^i(z^L,z^R)\text{ constant}} = -\frac{p}{1-p} \cdot \frac{f\left(z^L\right)}{f\left(z^R\right)} \cdot \frac{\eta\left(z^L;\bar{y}^i\right)}{\eta\left(z^R;\bar{y}^i\right)}$$

As the expression makes clear, indifference curves must be downward sloping. This should be intuitive; an increase in z^L is a concession from L- to R-faction judges that improves the utility of R-faction judges and lowers the utility of L-faction judges. To compensate, the R-faction judges must make a concession to L-faction judges, which requires z^R to fall.

---

[18]To see this, take some rule pair $\left(z_0^L, z_0^R\right)$. Since η is continuous and F admits a density, $v^i\left(z^L, z^R\right)$ is differentiable in the neighborhood of $\left(z_0^L, z_0^R\right)$, with: $\frac{\partial v^i}{\partial z^L} = -p\eta\left(z^L; \bar{y}^i\right) f\left(z^L\right) sgn\left(z^L - \bar{y}^i\right)$ and $\frac{\partial v^i}{\partial z^R} = -\left(1-p\right)\eta\left(z^R; \bar{y}^i\right) f\left(z^R\right) sgn\left(z^R - \bar{y}^i\right)$. Let $v_0^i = v^i\left(z_0^L, z_0^R\right)$ and $\phi^i\left(z^L, z^R\right) = v^i\left(z^L, z^R\right) - v_0^i$. Notice that $\phi^i\left(z_0^L, z_0^R\right) = 0$ and that $\frac{\partial \phi^i}{\partial z^R} = \frac{\partial v^i}{\partial z^R} \neq 0$ (except when $z^R = \bar{y}^i$). Then, by the implicit function theorem, there exists a function $z^R\left(z^L, v_0^i\right)$ s.t. $v^i\left(z^L, z^R\left(z^L; v_0^i\right)\right) = v_0^i$ in the neighborhood of $\left(z_0^L, z_0^R\right)$. This defines the indifference curve containing policy $\left(z_0^L, z_0^R\right)$. Moreover: $\frac{\partial z^R\left(z^L, v_0^i\right)}{\partial z^L} = -\frac{p}{1-p} \frac{f\left(z_0^L\right)}{f\left(z_0^R\right)} \frac{\eta\left(z_0^L; \bar{y}^R\right)}{\eta\left(z_0^R; \bar{y}^R\right)} \frac{sgn\left(z^L - \bar{y}^i\right)}{sgn\left(z^R - \bar{y}^i\right)}$.

Furthermore, the *IDID* property implies that, at any given point, the indifference curves of *L*-faction judges are less steep (than those of *R* faction judges) in the upper triangle, and steeper in the lower triangle.[19] Hence partial *stare decisis* regimes cannot be Pareto optima. By contrast, along the 45° line, all indifference curves have slope equal to $\frac{-p}{1-p}$. These represent the set of Pareto optimal rules.

The shaded area represents the set of legal rule-pairs that constitute Pareto improvements over Autarky. The logic of gains from trade are made clear by noting that from any rule pair $\left(z^L, z^R\right)$ with $z^L \neq z^R$, there emanates a 'lens' of Pareto improvements that widens in the direction of the 45° line. By the *IDID* property, the net utility gain from correctly deciding cases close to the ideal cut-point is small relative to the gain from correctly deciding cases far from the ideal cut-point. This implies that both types of judges value concessions by the other type to equivalent concessions[20] of their own, which is reflected in their indifference curves having different slopes. These differing valuations straight-forwardly produce opportunities for mutual gains from trade. Although partial *stare decisis* regimes often represent Pareto improvements over Autarky, the gains from trade only exhaust when judges implement complete *stare decisis* regimes.

We note, importantly, that whilst the Edgeworth Box suggests that the mechanism driving our results is trading over policies, in fact, what the judges are actually trading is case dispositions. Since judges never choose policies for the court, they do not possess the ability to trade policies. What they do possess is the ability (when recognized) to decide cases. Choosing among different policies amounts to the judges making trades about how they will dispose of various cases. (This is analogous to the more familiar framework where agents

---

[19]To see this, note that the slopes of the indifference curves differ only in the ratio $\frac{\eta\left(z^L; \bar{y}^i\right)}{\eta\left(z^R; \bar{y}^i\right)}$. Suppose $z^L < z^R$. The IDID property implies that $\eta\left(z^L; \bar{y}^L\right) < \eta\left(z^R; \bar{y}^L\right)$ and $\eta\left(z^L; \bar{y}^R\right) > \eta\left(z^R; \bar{y}^R\right)$. Hence, we have $\frac{\eta\left(z^L;\bar{y}^L\right)}{\eta\left(z^R;\bar{y}^L\right)} < 1 < \frac{\eta\left(z^L;\bar{y}^R\right)}{\eta\left(z^R;\bar{y}^R\right)}$.

[20]By equivalent concessions, we mean that the expected fraction of cases implicated is the same. Suppose $z^L < z^{L'} < z^{R'} < z^R$. Then the concessions by L from z^L to $z^{L'}$, and by R from z^R to $z^{R'}$ are equivalent provided that $p\left[F\left(z^{L'}\right) - F\left(z^L\right)\right] = (1-p)\left[F\left(z^R\right) - F\left(z^{R'}\right)\right]$.

choose amongst bundles by trading quantities of goods.) Although the judges' decisions are over options in policy space, the import of those policy decisions is determined by their consequence for outcomes in case space. This connection between the case space and policy space is crucial to understanding this paper's main insights.

Finally, we note that Lemmas 1 and 2 together imply that the incentives for the judges in the adjudication game are strategically equivalent to a (continuous action) Prisoners' Dilemma. Although there are potential gains from cooperation, the individual incentives for each judge cause them to optimally choose a Pareto inferior regime, in equilibrium. In the next section, we investigate the possibility that repeated interaction between the players may provide incentives for cooperation.

4   Sustainability of *Stare Decisis*

The gains from trade offered by complete stare decisis make it an attractive regime for judges. But judges have no mechanism to commit themselves to such a regime. Confronted with a case she must decide incorrectly under *stare decisis*, a judge may be sorely tempted to defect and decide the case correctly according to her own lights. We thus seek to identify conditions that will sustain a practice of *stare decisis* as an equilibrium in an infinitely repeated game between the judges on the bench.

4.1   Sustainable Stare Decisis

It is a well known result that non-Nash outcomes can be sustained in a repeated game if the players employ strategies that punish deviations from the designated action profile. Whilst a variety of potential punishment strategies exist, in this section, we study a focal strategy:

grim trigger Nash reversion.[21] Let z^C be a compromise rule, and let $H_t\left(z^C\right)$ be the set of histories up to time t in which all previous actions were consistent with judges respecting the complete *stare decisis* regime $\left(z^L, z^R\right) = \left(z^C, z^C\right)$. Consider the pair of strategies $\left(\sigma^L, \sigma^R\right)$ defined by:

$$
\sigma^i\left(h_t, x\right) = \begin{cases} r\left(x; z^C\right) & \text{if } h_t \in H_t\left(z^C\right) \\[2mm] r\left(x; \bar{y}^i\right) & \text{otherwise} \end{cases}
$$

Each judge respects the complete *stare decisis* regime provided that it had been respected in all previous periods, and reverts to autarky otherwise. This pair of strategies can be sustained as an equilibrium if, but only if, for any case that a judge is recognized to adjudicate, adherence to *stare decisis* yields a larger discounted lifetime utility stream than deviating to autarky. A critical point to note is that the greatest temptation to defect from *stare decisis* occurs when the case confronting the ruling judge is located precisely at the compromise policy z^C, for such a case is the most-distant (and hence costly) case the judge is obliged to decide incorrectly under *stare decisis.* Hence, if the ruling judge would adhere to *stare decisis* when $x = z^C$, then she will adhere to *stare decisis* for any other case as well. In this most demanding case, judge i will adhere to *stare decisis* provided:

$$
l\left(z^C; \bar{y}^i\right) + \frac{\delta}{1-\delta} v^i\left(z^C, z^C\right) \geq g\left(z^C; \bar{y}^i\right) + \frac{\delta}{1-\delta} v^i\left(y^L, y^R\right)
$$

which implies:

$$
\delta \geq \frac{\eta\left(z^C; \bar{y}^i\right)}{\eta\left(z^C; \bar{y}^i\right) + \left[v^i\left(z^C, z^C\right) - v^i\left(y^L, y^R\right)\right]}
$$

.

---

[21]Of course, other types of punishments strategies exist, and the analysis that follows can be easily modified to accommodate different punishments. Importantly, we acknowledge that punishments exist that are more severe than Nash reversion. However, we also note that such punishments may not necessarily be credible if δ is sufficiently small. By contrast, since Autarky is a Nash equilibrium of the stage game, the threat of Nash reversion is always credible. We also note that cooperation may be sustained by less severe punishment strategies, especially when δ is large.

Deviation to autarky improves the deliberating judge's short run utility (since she can now dispose of cases according to her ideal rule), but causes expected utility to fall in all future periods, since reversion to autarky represents a Pareto-deterioration. If the judge values future cases sufficiently highly, then the long run losses will outweigh the short-run gains, and cooperation will be sustainable.

For each $i \in \{L, R\}$, let

$$\delta^i \left(z^C \right) = \frac{\eta \left(z^C; \bar{y}^i \right)}{\eta \left(z^C; \bar{y}^i \right) + \left[v^i \left(z^C, z^C \right) - v^i \left(y^L, y^R \right) \right]}$$

and let $\delta \left(z^C \right) = \max \left\{ \delta^L \left(z^C \right), \delta^R \left(z^C \right) \right\}$. Since to sustain an equilibrium, *all* judges must be incentivized to cooperate, we need that $\delta \geq \delta \left(z^C \right)$.

Proposition 1. *A complete stare decisis regime at compromise threshold z^C can be sustained as an equilibrium of a repeated game, provided $\delta \geq \delta \left(z^C \right)$. Moreover:*

1. *there exist $\underline{z}, \bar{z} \in \left(\bar{y}^L, \bar{y}^R \right)$ with $\underline{z} \leq \bar{z}$, such that $\delta \left(z^C \right) > 1$ whenever $z^C < \underline{z}$ or $z^C > \bar{z}$.*

2. *there exists a unique $\delta^* \in (0, 1)$ and $z^* \in (\underline{z}, \bar{z})$, such that $\delta^* = \delta \left(z^* \right)$ and $\delta \left(z^C \right) \geq \delta^*$ for all z^C. Further $\delta \left(z^C \right)$ is decreasing for $z^C < z^*$ and increasing for $z^C > z^*$.*

3. *the correspondence $Z \left(\delta \right) = \left\{ z^C | \delta \geq \delta \left(z^C \right) \right\}$ is monotone in the sense that $\delta' > \delta$ implies $Z \left(\delta \right) \subset Z \left(\delta' \right)$.*

Several points are worth noting. First, although they are Pareto optimal, not all complete *stare decisis* regimes are sustainable. If the compromise legal rule is skewed too far towards the ideal legal rule of either faction, then cooperation breaks down. The thresholds $\underline{z}$ and $\bar{z}$ bound the set of potentially sustainable compromise policies. These correspond to the boundaries of the shaded region in Figure 1 along the 45° line; they are the subset of Pareto optimal rules that are Pareto improvements over autarky.

Figure 2: Sustainable *Stare Decisis*.
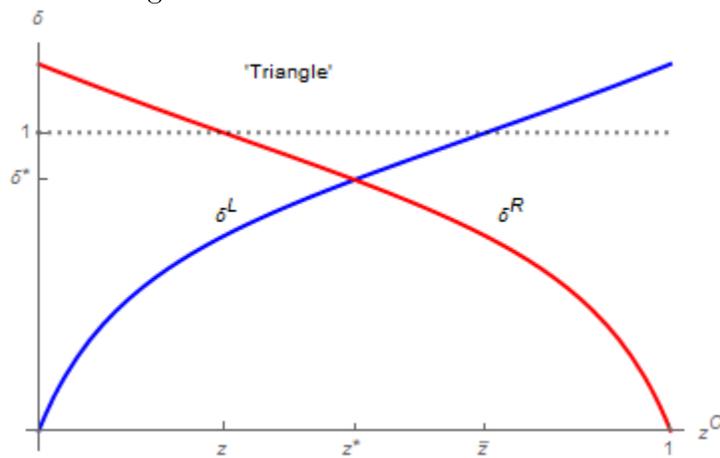
The horizontal axis represents compromise policies z^C, the vertical axis represents the discount factor δ. The upward sloping curve is δ^L, the critical discount factor for L judges. The downward sloping curve is δ^R. The discount factor cannot be larger than 1. Complete *stare decisis* is sustainable when the discount factor is larger than both critical values. The figure assumes the environment in Example 1, with $\bar{y}^L = 0$, $\bar{y}^R = 1$, $\varepsilon = \frac{2}{3}$, and $p = \frac{1}{2}$.

Second even among the set of Pareto improving compromises ($z^C \in (\underline{z}, \bar{z})$), these policies are only sustainable if the discount factor is sufficiently large. The function $\delta\left(z^C\right)$ determines how large the discount factor must be before a compromise at z^C is sustainable. Obviously, if $\delta\left(z^C\right) > 1$, then it is impossible to sustain cooperation around the cutoff point z^C, whereas if $\delta\left(z^C\right) = 0$, then such a compromise is guaranteed to be sustainable. As part (2) of the proposition shows, $\delta\left(z^C\right)$ is always strictly greater than zero, and so if judges are sufficiently impatient, then *none* of the Pareto optimal rule-pairs are implementable in equilibrium. Moreover, δ^* is the minimum discount factor needed to sustain cooperation at *some* compromise cut-point.

Third, as the discount factor increases, the set of compromise policies that can be sustained grows. We see this in Figure 2, which plots the set of compromise policies that can be sustained for each value of δ. The sustainable set is the triangle-like region bounded by the line $\delta = 1$ and the curves $\delta^L\left(z^C\right)$ and $\delta^R\left(z^C\right)$. The bottom tip of the triangle represents the pair (z^*, δ^*). Monotonicity is evidenced by the triangle 'widening' as δ increases.

26

The failure to sustain Pareto optimal policies stems from a problem of commitment. Although both types of judges favor compromise policies *ex ante*, once chosen, the presiding judge has a short-run incentive to defect and dispose of cases according to her ideal legal rule. Cooperation in the short run is sustained by the threat of a breakdown of cooperation in the future. Hence, cooperation requires that the short-run cost of adhering to the agreement is smaller than the discounted long-run value of gains from trade. If δ is low, then cooperation is not sustainable as the future gains from trade are insufficient to compensate the presiding judge for short run losses.

We can interpret the size of the discount factor in two ways. On the one hand, if cases arrive at a fixed rate, a high δ implies that judges are relatively far-sighted and place sufficient value on the disposition of future cases. Under this interpretation, *stare decisis* should be more likely to be in effect in high salience issue spaces, where judges care deeply about the future impact of policy choices. On the other hand, if the case arrival rate is not fixed, a high δ may imply that the next case will arrive relatively soon into the future, and so the gains from defection will be relatively short-lived. By contrast, a low δ may imply that a long period of time will pass before a judge from the other faction has the opportunity to reverse the current judge's decision, and so the gains from defection may persist for a long time. Under this interpretation, our model predicts that compromise policies are more likely to arise in areas of the law that are frequently litigated, than in areas where controversies arise rarely.

4.2 Comparative Statics

In the previous sub-section, we characterized the set of compromise policies z^C that can be sustained in equilibrium, given the discount factor δ. In this sub-section, we are interested in how these sustainable compromises are affected by heterogeneity and polarization. Doing so will enable us to highlight empirical predictions of the model. In analyzing the comparative

statics, we will be concerned with how the underlying parameters affect: (i) the functions δ^L and δ^R, (ii) the compromise policy z^* that is implementable over the broadest set of discount factors (abusing terminology, the 'most frequent' compromise policy), (iii) the lowest discount rate at which some compromise is feasible δ^*, and (iv) the highest and lowest compromise policies ($\bar{z}$ and $\underline{z}$, resp.) that can be sustained.

4.2.1 Heterogeneity and Sustainability

Recall, heterogeneity is a measure of the distribution of power between L and R judges on the bench. In our model, this is captured by the parameter p. Heterogeneity is highest when $p = \frac{1}{2}$, so that both factions are equally powerful. Heterogeneity has the following implications for equilibrium policy:

Lemma 3. *Following an increase in p:*

1. *L-faction (resp. R-faction) judges become less (resp. more) amenable to compromise. Formally, $\delta^L\left(z^C\right)$ (resp. $\delta^R\left(z^C\right)$) is increasing (resp. decreasing) in p.*

2. *The 'most frequent' compromise policy z^* moves closer to L's ideal (z^* is decreasing in p). The effect on the lowest discount factor $\delta^* = \delta\left(z^*\right)$ is ambiguous.*

3. *The highest and lowest sustainable policies both move closer to L's ideal ($\bar{z}$ and $\underline{z}$ are both decreasing in p), although the effect on the range of compromise policies ($\bar{z} - \underline{z}$) is ambiguous.*

Lemma 3 makes several points. First, it shows that, as the L faction gains more power, L-faction judges will be less amenable to compromise whilst R-faction judges become more amenable. In the left panel of Figure 3 below, this manifests as an upward shift in the δ^L curve and a downward shift in the δ^R curve. There is a set of compromise policies that L faction

judges would have previously supported, but which they no longer will. Similarly, there is a set of compromise policies that R faction judges would have previously not supported, but which they now will. As judges become more politically powerful, they become less amenable to compromise. The intuition is straight-forward; as p increases, the cost of autarky falls for L-faction judges and rises for R-faction judges. The implications for the benefits from compromise are obvious.

Second, the 'most frequent' compromise policy z^* (i.e. the bottom tip of the triangle in Figure 2) shifts towards the L-faction's ideal. Similarly, the highest and lowest sustainable compromise policies ($\bar{z}$ and $\underline{z}$, respectively) both shift towards the L-faction's ideal. These results follow immediately from the preceding insight that L-faction judges can be more demanding, whilst R-faction judges will be less so. By contrast, the effect on the range of discount factors that can sustain a compromise is ambiguous. If the degree to which L faction judges have become more demanding exceeds the degree to which R faction judges have become more amenable, then compromises will be less likely, and vice versa. The range of compromise policies ($\bar{z} - \underline{z}$) that can be sustained as an equilibrium is similarly ambiguous.

Figure 3 depicts these comparative statics for our working example. The left panel shows how the curves δ^L and δ^R shift in response to an increase in p, and the implications for the sorts of compromise policies that can be sustained. The right panel shows how the highest, most frequent and lowest compromise policies change as p increase from 0 to 1.

4.2.2 Polarization and Sustainability

In our model, polarization ρ is measured as the fraction of cases in the region of conflict, $\rho = F\left(\bar{y}^R\right) - F\left(\bar{y}^L\right)$.[22] As ρ increases from 0 to 1, so does the fraction of cases over which

---

[22]At first blush, this definition might seem odd. Why not simply define polarization as the absolute distance between the ideal cutpoints $\bar{y}^R - \bar{y}^L$? But this ignores that the case-space, as defined so far, is

Figure 3: The Effect of Heterogeneity on the Set of Sustainable Compromise Policies.
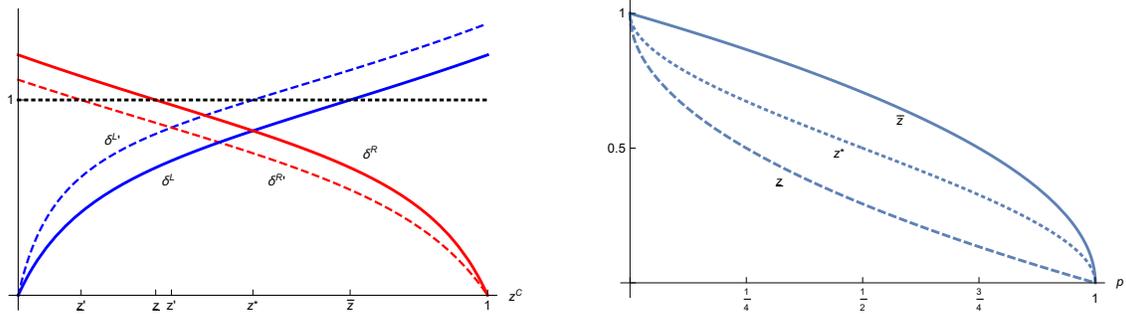
The figure assumes the environment in Example 1, with $\bar{y}^L = 0$, $\bar{y}^R = 1$, and $\varepsilon = \frac{2}{3}$. In the left panel, the solid lines represent δ^L and δ^R when $p = \frac{1}{2}$, whilst the dashed lines represent the same functions when $p = \frac{3}{4}$. We see that δ^L shifts up whilst δ^R shifts down, and consequently $\underline{z}, z^*$ and $\bar{z}$ all decrease. (Note, in this particular example, z^* and $\bar{z}'$ happen to coincide.) The right panel depicts $\underline{z}, z^*$ and $\bar{z}$ as p increases from 0 to 1.

the judges will disagree, and hence, the likelihood of conflict.

Let $F(x; [a, b])$ denote the conditional distribution of X, given that $X \in [a, b]$. To analyze the effect of a change in polarization, note that judicial preferences over legal rules can be written:

$$v^i\left(z^L, z^R\right) = E\left[g\left(\theta\right)\right] - \rho \left\{ p \left| \int_{y^i}^{z^L} \eta\left(x; \bar{y}^i\right) dF\left(x; \left[\bar{y}^L, \bar{y}^R\right]\right) \right| + (1-p) \left| \int_{y^i}^{z^R} \eta\left(x; y^i\right) dF\left(x; \left[\bar{y}^L, \bar{y}^R\right]\right) \right| \right\}$$

As before, the relevant component of utility is the second term, which is the product of the degree of polarization and expressions that only depend on the conditional distribution of cases within the region of conflict. In performing the comparative static analysis, we consider changes in the case-distribution that cause the degree of polarization to change, holding fixed the conditional distribution in the region of conflict. Hence, changing our measure of polarization changes the likelihood that the judges will be in conflict, but does not change the relative likelihood of certain controversial cases arising over others.[23]

---

endowed neither with a natural 0 case, nor with a standard unit of measurement. Hence, we have two degrees of freedom, and it is without loss of generality to fix $\bar{y}^L$ and $\bar{y}^R$, and to scale all other cases accordingly. (In our example, we specify $\bar{y}^L = 0$ and $\bar{y}^R = 1$.) Hence, $\bar{y}^L$ and $\bar{y}^R$ will be invariant in our model. However, we can think of these cutpoints as being 'closer' or 'further' from each other, relative to the set of cases that will likely arise, given the case generating process. This motivates our chosen measure for polarization.

[23]It should be clear that changing the conditional distribution will change the relative value of compromise

Lemma 4. *Following an increase in polarization $\rho = F\left(\bar{y}^R\right) - F\left(\bar{y}^L\right)$ (supposing that the conditional distribution $F_{X \in [\bar{y}^L, \bar{y}^R]}$ is unchanged):*

1. *Both L and R faction judges become more amenable to compromise. Formally, $\delta^L\left(z^C\right)$ and $\delta^R\left(z^C\right)$ are both decreasing in ρ, for $z^C \in [\underline{z}, \bar{z}]$.*

2. *The 'most frequent' policy z^* and the highest and lowest sustainable policies ($\bar{z}$ and $\underline{z}$, resp.) are unchanged.*

3. *The lowest discount factor at which stare decisis is sustainable δ^* decreases.*

Polarization has a remarkable effect in the model: it increases the gains from trade and hence expands the set of compromise policy/discount-factor pairs $\left(z^C, \delta\right)$ that are consistent with *stare decisis*. The ' triangle' becomes larger. The result is somewhat counter-intuitive; one might have expected that the more the factions disagree, the less likely they would be amenable to compromise. But this ignores the gains from trade logic that underpins the model. As polarization increases, the expected losses from autarky become larger, since more cases will now be decided 'incorrectly' if the presiding judge is from the opposing faction. This implies that the gains from trade have become larger, making defection costlier, and cooperation easier to sustain.

Figure 4 depicts the effect of a decrease in polarization in the case of our working example. Given the uniform distribution, polarization in the working example is given by $\rho = \frac{1}{2\varepsilon}$. An increase in polarization causes both the δ^L and δ^R curves to shift downward. This implies that more compromise policies become sustainable for any given $\delta < 1$. Moreover, both curves shift down by the 'same amount', so that the policy z^* at which they intersect (i.e. the 'most-frequent' compromise policy) is unchanged.

---

to L and R faction judges, independent of the distance between their ideal points.

Figure 4: The Effect of Polarization on the Set of Sustainable Compromise Policies.
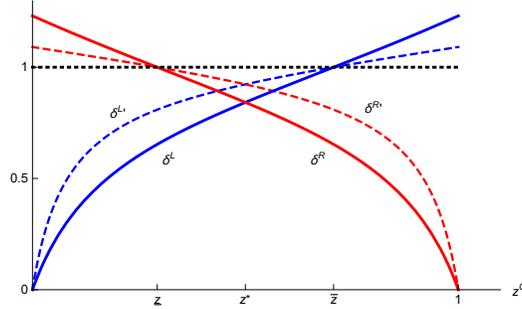
The figure assumes the environment in Example 1, with $\bar{y}^L = 0$, $\bar{y}^R = 1$, and $p = \frac{1}{2}$. The solid lines represent δ^L and δ^R when $\rho = \frac{3}{4}$ (i.e. $\varepsilon = \frac{2}{3}$), whilst the dashed lines represent the same functions when $\rho = \frac{1}{3}$ (i.e. $\varepsilon = \frac{3}{2}$).

5   Partial Stare Decisis

In the previous section, we showed that problems of commitment explain the inability of judges to sustain complete *stare decisis* regimes, even though they are efficient. In this section, we show that judges may credibly commit to partial compromises even when complete compromises are not feasible. Moreover, the nature of this partial compromise resembles the operation of a standard – judges from both factions agree to dispose of 'relatively extreme' cases in a consistent manner, but retain the discretion to dispose of 'relatively moderate' cases according to their ideal. The law is partly-settled-partly-contested, and this regime is persistent.

To build intuition for this result, suppose, beginning from autarky, the judges progressively move their thresholds towards the middle until the thresholds meet. By Lemma 2, we know that whenever the thresholds do not coincide, there are further gains from trade to be exploited. However, the lion's share of the gains from trade are actually realized at the start. To see this, consider the L faction judges; they lose little in the original compromise since the disutility from incorrectly deciding cases close to their threshold is small, whilst the utility gain from having cases far from their threshold (close to R's threshold) decided 'correctly' is large. As the judges progressively compromise, the size of the gains fall relative

32

to the losses, and so the net gains from trade, whilst positive, decrease.

Recall, also, that sustaining cooperation becomes harder as the compromise threshold moves further away from the judge's ideal. Hence, as judges compromise more, the marginal disutility from compromising in the stage game increases, whilst the marginal utility gain in the continuation game decreases. It may be that, on net, the net gains (adding costs and benefits in the stage and continuation game) become negative before the thresholds meet. I.e. it may be more desirable to sustain a partial compromise than the complete one.

These intuitions are formalized in the following proposition that generalizes the insights from the previous section. Let $\left(z^L, z^R\right)$ be a pair of legal rules that constitute a Pareto improvement over autarky (i.e. which are contained in the shaded lens in Figure 1). For each $i \in \{L, R\}$, let

$$\delta^i \left(z^L, z^R\right) = \frac{\eta \left(z^i; \bar{y}^i\right)}{\eta \left(z^i; \bar{y}^i\right) + \left[v^i \left(z^L, z^R\right) - v^i \left(\bar{y}^L, \bar{y}^R\right)\right]}$$

and let $\delta \left(z^L, z^R\right) = \max \left\{\delta^L \left(z^L, z^R\right), \delta^R \left(z^L, z^R\right)\right\}$. Then the partial *stare decisis* regime $\left(z^L, z^R\right)$ is sustainable provided that $\delta \geq \delta \left(z^L, z^R\right)$.

Proposition 2. *The following are true:*

1. *the correspondence $Z \left(\delta\right) = \left\{\left(z^L, z^R\right) \mid \delta \geq \delta \left(z^L, z^R\right)\right\}$ is monotone in the sense that $\delta' > \delta$ implies $Z \left(\delta\right) \subset Z \left(\delta'\right)$.*

2. *autarky is always sustainable (i.e. $\left(\bar{y}^L, \bar{y}^R\right) \in Z \left(\delta\right)$ for all $\delta \in [0, 1]$).*

3. *autarky is the only sustainable regime for δ small enough. Formally, there exists $\delta^{**} \in [0, \delta^*)$ such that $Z \left(\delta\right) = \left\{\left(\bar{y}^L, \bar{y}^R\right)\right\}$ for all $\delta \leq \delta^{**}$. Furthermore $\delta^{**} = 0$ iff $\frac{\partial \eta \left(x; \bar{y}^i\right)}{\partial x}\big|_{x=\bar{y}^i} = 0$ for each i.*

4. *for intermediate values of δ, partial stare decisis regimes are sustainable, although*

complete stare decisis regimes are not. Formally, if $\delta \in (\delta^{**}, \delta^{*})$, then there exists $\left(z^L, z^R\right) \neq \left(\bar{y}^L, \bar{y}^R\right) \in Z\left(\delta\right)$, and if $z^L = z^R$, then $\left(z^L, z^R\right) \notin Z\left(\delta\right)$.

Proposition 2 bears many similarities to Proposition 1, which is to be expected; the same logic underpins both claims. Two differences are worth noting. The second part of Proposition 2 has no analogue in Proposition 1, but it is not interestingly new. It simply reflects a result we have previously stated, that since autarky is a Nash equilibrium of the stage game, it must be an equilibrium in the repeated game. The fourth part is substantially different; it states that there are a range of environments over which partial *stare decisis* regimes may be sustainable whereas complete *stare decisis* regimes cannot.

Proposition 2 also shows that, if the judges are sufficiently present biased, then no compromise may be possible at all, and the only feasible regime is autarky.[24]

The intuition for these results can be seen in Figure 5. In each panel, the thin outer curves represent the locus of policies over which the L and R-faction judges (respectively) are indifferent to autarky. (These curves were represented in Figure 1 as well.) The lens enclosed by these indifference curves represents the set of Pareto improvements over autarky. These are also the set of sustainable (partial) compromises when $\delta = 1$ (since in this case, judges do not care about short run losses). The thick inner curves represent the locus of policies which are just sustainable (for the L and R-faction judges, respectively) given some discount factor $\delta < 1$. In the left hand panel, $\delta > \delta^{*}$, and so some complete *stare decisis* compromises are sustainable. In the right hand panel $\delta < \delta^{*}$, and so none of the complete *stare decisis* regimes are feasible, but some partial regimes are. The set of policies contained within the smaller lens is precisely $Z\left(\delta\right)$.

---

[24]To understand the condition on the derivative of η, note that since η is increasing, starting from autarky, an ε-compromise by both factions of judges generates a first order gain-from-trade; future utility from the compromise increases in the first order. If $\eta' > 0$ at autarky, then the immediate utility loss from compromise is also of the first order. When judges are sufficiently impatient, the latter will dominate the former. By contrast, if $\eta' = 0$ at autarky, then the immediate utility loss from compromise is of the second order. The future gains from trade will always dominate, and so a sustainable compromise always exists.

Figure 5: Sustainable Legal Rules and the Discount Factor
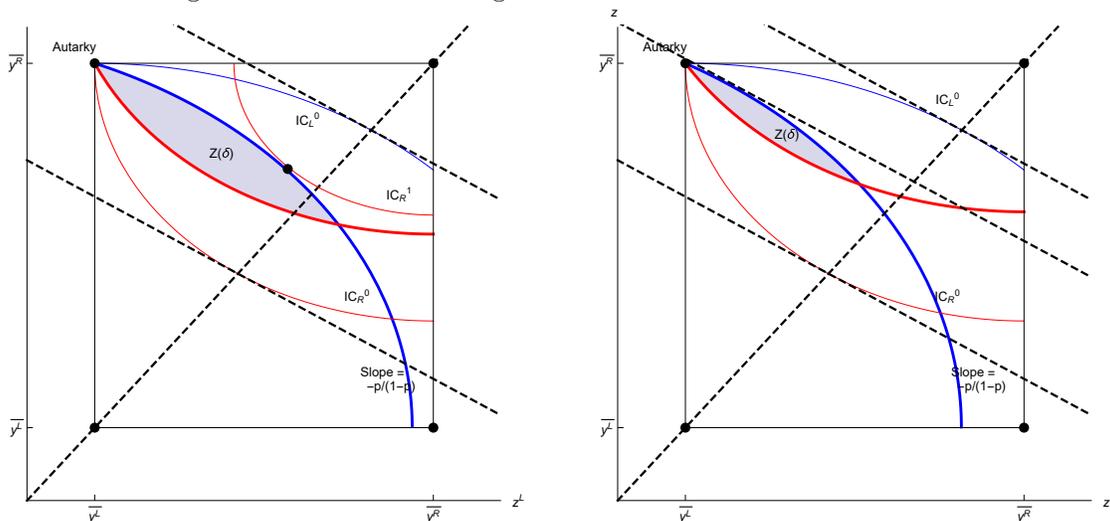
The figure assumes the environment in Example 1, with $p = \frac{1}{3}$. In the left panel, $\delta = 0.88$, whilst in the right panel, $\delta = 0.83$.

As δ decreases the inner curves move 'towards each other', causing the set of sustainable policies to shrink. This is the monotonicity property in part 1 of the proposition. The δ^* defined in Proposition 1, is the δ for which the two curves intersect precisely on the 45° line. In both panels, at Autarky, the inner curves expand 'away from one another'. However, since these curves also move towards one another as δ decreases, there may be some δ small enough, such that the curves become tangent to one another at Autarky. This defines δ^{**}. For $\delta \leq \delta^{**}$, the two curves no longer enclose a lens of policies. Autarky is the unique sustainable equilibrium. (However, if $\eta'\left(\bar{y}^L; \bar{y}^L\right) = 0$, then L's curve must be horizontal at autarky for any δ, and if $\eta'\left(\bar{y}^R; \bar{y}^R\right) = 0$, then R's curve must be vertical at autarky for any δ. In either case, it is impossible for the curves to ever be tangent, and so a lens is enclosed for all $\delta > 0$.)

5.1 Equilibrium Selection

Our analysis so far has identified multiple rule pairs $\left(z^L, z^R\right)$ that are equilibrium consistent. Amongst these, we might think that some regimes are more likely to arise than others.

Although equilibrium selection is of significant interest, it is beyond the scope of this paper to explore that issue in depth. In this section, we simply document the possibility that, under a simple and not implausible selection procedure, judges will routinely converge on partial *stare decisis* regimes, even when complete *stare decisis* regimes are feasible. This insight highlights that standards need not only arise in contexts where it is infeasible to implement a coherent rule. Instead, standards may prevail in a broader set of contexts, including some where the judges might have otherwise settled on a common legal rule.

Consider the following simple equilibrium selection mechanism: prior to the adjudication game proper, one of the factions is selected to propose a compromise policy from the sustainable set $Z(\delta)$. If the proposal is accepted by the other faction, then the adjudication game begins and the judges coordinate on the proposed policy. If the proposal is rejected, then the judges expect that autarky will prevail. This selection procedure coincides with the agenda setter framework of Romer and Rosenthal (1978) or the one-period bargaining protocol in Baron and Ferejohn (1989). Naturally, the procedure confers a significant benefit on the recognized faction. A natural interpretation is that the agenda setting faction is the faction that is ascendant during the period in which some doctrinally important case arises. We re-iterate that we are not arguing that this selection procedure is ideal or most plausible. We simply note that it is a not unreasonable procedure, and we explore its implications.

Proposition 3. *Suppose $\delta < 1$, and one faction is free to select any equilibrium to implement. The agenda setter will choose a partial stare decisis regime, unless $\delta < \delta^{**}$, in which case the only feasible regime is Autarky.*

Proposition 3 shows that if the equilibrium selection mechanism sufficiently privileges one faction over the other, it will select a Pareto inferior partial *stare decisis* regime, even when Pareto optimal regimes are sustainable.[25] This seems counter-intuitive – we might have

---

[25]Of course, the chosen policy will be constrained Pareto optimal, given the sustainability constraint. If not, the agenda setter could improve her utility further.

conjectured that the agenda setter would choose the most favorable sustainable Pareto optimum. To build intuition for this result, note that judge i's policy utility from an arbitrary rule pair $\left(z^L, z^R\right)$ can be decomposed as the difference of two terms: the policy utility associated with the complete *stare decisis* regime implementing the expected threshold $E\left[z\right] = pz^L + \left(1 - p\right) z^R$, less a term proportional to the variance of the rule.[26] Each judge prefers a complete *stare decisis* regime (since the variance term is zero) to any other rule-pair that induces the same expected rule, *ceteris paribus*. But as we showed in Proposition 2, there may be partial *stare decisis* regimes that are sustainable, even when the corresponding complete *stare decisis* regime is not. Given this constraint, the proposing faction faces a trade-off between pulling the expected threshold closer to their ideal (which improves utility along the first dimension) and increasing the variance of the rule (which decreases utility along the second dimension). At the optimum, the proposer is willing to introduce *some* variance into the rule, in order to bring the implemented rule closer to her ideal, in expectation. Hence, the proposing faction chooses to not realize all potential gains from trade. Even when a perfect compromise is available, we may expect to observe legal doctrines that are partially settled and partially contested.[27]

This is made apparent in the left panel of Figure 5. The figure depicts the case when the R faction proposes the agenda. As before, the shaded region represents the set of sustainable policies, which in this case includes some complete *stare decisis* policies. The proposer will choose the policy that puts him on the highest achievable indifference curve $\left(IC_R^1\right)$ given the sustainability constraint. The slopes of the R faction judges' indifference curves, and the boundary of $Z\left(\delta\right)$ make clear that the optimum is incompatible with perfect *perfect stare decisis*.

---

[26]Formally, we take a second order Taylor approximation of $v^i\left(z^L, z^R\right)$ centered at $E\left[z\right] = pz^L + \left(1 - p\right) z^R$. This gives $v^i\left(z^L, z^R\right) \approx v^i\left(E\left[z\right], E\left[z\right]\right) - \psi^i\left(E\left[z\right]\right) Var\left(z\right)$, where $\psi^i\left(z\right) = \frac{d}{dz}\left(\eta\left(z; \bar{y}^i\right) f\left(z\right)\right)$ and $Var\left(z\right) = p\left(z^L - E\left[z\right]\right)^2 + \left(1 - p\right)\left(z^R - E\left[z\right]\right)^2$.

[27]We acknowledge that other procedures will select complete *stare decisis* regimes whenever available. For example, if we took the polar opposite selection mechanism, where both factions engaged in Nash Bargaining to determining the equilibrium policy, then complete *stare decisis* regimes would be selected.

5.2 Partial *Stare Decisis* in Practice

The notion of partial *stare decisis* is novel in our analysis. These rules have been ignored in both the formal and legal literatures on judicial behavior. Nevertheless these rules play an important role in our analysis. Over a range of δ, the only, non-autarkic equilibria are partial *stare decisis* rules. Moreover, we saw that, in the agenda setter framework, a partial *stare decisis* rule that does not exhaust the potential gains from trade is the equilibrium selected even when a complete *stare decisis* rule would be sustainable and would exhaust the gains from trade.

Recall that under a rule (z^L, z^R) (with $z^L \neq z^R$), the bench decides cases, for $x < z^R$, as the right faction would. Similarly, it decides cases $x > z^R$ as the L faction would. In the interval (z^L, z^R), we have autarky. This structure implies that a body of law governed by a rule of partial *stare decisis* may look incoherent over the interval (z^L, z^R). Moreover the larger this interval, the more incoherent the body of law will appear. With this incoherence as a marker, we might identify a number of highly criticized doctrines as bodies of law governed by partial *stare decisis*. The fact that no clear complete rule has emerged should not conceal the fact that judges with distinct jurisprudential views on the ideal rule have managed to reduce the extent of their disagreement and to adhere partially to *stare decisis*.

Consider, for example, search and seizure law in the United States. The Constitution protects individuals against "unreasonable searches and seizures" of their persons and property. Challenges to warrantless searches and seizures are frequent and the law is generally considered incoherent.[28] There are clear cases, at both extremes, on which the left and the right are agreed. Between these extremes, the decisions seem chaotic and incoherent. That incoherence is consistent with partial *stare decisis* as, between the extremes, decisions will be rendered by the judge designated to decide the case.

---

[28]Wasserstrom and Seidman (1988), for example, cite critics on both the right and the left.

In terms of our model, one might identify the one-dimensional case space as the "degree of intrusiveness of the search". At the extreme intrusive end of the spectrum lie searches incident to unwarranted intrusions into the individual's home; on the other end of the spectrum lie "searches" of, say, cars that uncover drugs in "plain view." Admittedly, the area in between the endpoints of the policy may be vast but the courts can be understood as following a rule of partial *stare decisis*.

As a second, somewhat stylized, example, consider the rules governing which terms constitute the contract. This issue arises often in on-line transactions when a consumer assents not only to the core terms of the contract (such as price and physical product description) as well as to "additional terms" that are more or less easily accessed. In other cases, these additional terms are only revealed after the product is delivered ("buy now, terms later" agreements). These questions arise frequently both in on-line and physical transactions. For purposes of the model, consider a one-dimensional case space that one might call "conspicuousness". Conspicuousness refers to prominence of a notification of the existence of additional post-formation terms and the ease with which they may be accessed. Courts generally agree that low conspicuousness implies unenforceability of the "hidden" terms while high conspicuousness implies enforceability.[29] For intermediate degrees of conspicuousness, consumer-oriented judges decide one way – typically that the post-formation terms are not part of the contract – while business-oriented judges decide the other way.

The use of rules of partial *stare decisis* thus identify an underlying structural coherence to bodies of law that superficially appear to be chaotic and confused.[30]

---

[29]Restatement of Consumer Contracts § 2 (Am. Law Inst., Discussion Draft, 2017). We thank Florencia Marotta-Wurgler for directing us to this doctrine and source.

[30]There are other potential explanations. The law might, of course, simply be confused. Or the law might be evolving with the judges clarifying the law over time. On this latter account, we should observe the size of the interval (z_L, z_R) shrinking over time.

6   Conclusion

In this paper, we studied the political logic of *stare decisis*. We showed that, if judges preferences satisfy a simple and reasonable property – the *increasing differences in dispositional value (IDID)* property – then heterogeneous judges may achieve gains-from-trade by implementing a single, consistent legal rule. However, commitment problems may prevent them from realizing these gains. We showed that judges may credibly commit to deciding cases as if in accordance with a common rule, if they are sufficiently far sighted. We associated far-sightedness with issue-spaces that are highly salient or over which litigation is frequent. We further showed, counter-intuitively, that commitment is more likely when the bench is ideologically polarized, since the gains from trade are largest on such benches. Our model, thus, makes several empirical predictions that we can take to the data.

An important and novel contribution of our paper is in identifying and providing a theoretical basis for the notion of partial *stare decisis*. We showed that partial *stare decisis* equilibria have a structure akin to legal standards (rather than rules) and may be sustained even in cases when judges cannot credibly commit to applying coherent rules. Additionally, we argued that partial *stare decisis* regimes may prevail even in circumstances when complete *stare decisis* equilibria were feasible. Hence, we argue that the scope for partial *stare decisis* is potentially quite significant. Our model demonstrates that the notions of complete and partial *stare decisis* are mutually consistent with rational judicial decision making, and provides a framework to understand why legal doctrines are characterized by fully settled law in some instances, and are left unsettled in part, in others.

Our formalization of *stare decisis*, and especially the partial variety, demonstrates the importance of the case-space approach to modeling judicial decision making. Indeed, much of the interesting action in our paper stems from the sequential and unilateral nature of decision making, which makes it challenging for judges to credibly commit to applying compromise rules. Case-space provides the natural framework within which to analyze the problem of

commitment and consistency, since, as we discussed in the introduction, these concerns are inextricably linked to cases and case dispositions. Indeed, the novelty of partial *stare decisis,* the account of why courts only apply precedent over a subset of cases, required the technology of the case-space framework.

Although our model focuses on the judicial context, the insights have a broader reach. Adjudication is a paradigmatic example of a task in which the decision maker must sort cases or events into categories. Preferences over dispositions takes this task seriously and attributes to the decision maker preferences that reward her for correctly sorting the cases. Many bureaucrats face similar tasks – e.g., insurance agents declining coverage or immigration officials granting asylum. Understanding the role that various assumptions, including IDID, on such dispositional preferences is thus apt to have wide application.

Finally, we suggest some avenues for future research. We have assumed that the case-generating process does not vary with the prevailing rule of decision on the court. This assumption is common in models of courts but highly likely to be counterfactual. Litigants are apt to settle "easy" cases and litigate "hard" ones. Case difficulty will depend on the prevailing rule of decision. Thus we might expect that, under autarky, most litigated cases will fall within the conflict zone and very few outside it. Under a rule of complete *stare decisis,* we might expect most litigated cases to fall near the cut-off point z^C of the compromise legal rule while under a rule of partial *stare decisis*, litigated cases might fall primarily within the interval $[z^L, z^R]$. Rule-dependent case-generating processes will thus alter the size of the gains from trade available to the two factions. Further work might determine the extent to which this undermines or supports a practice of *stare decisis.*

We explicitly model a structure similar to that of an intermediate court of appeals in which cases are decided by randomly drawn panels of judges. But we think a similar mechanism might be at play on courts such as the Supreme Court of the United States that sit *en banc.* Author-influence models of these courts, (e.g. Lax and Cameron (2007)), suggest that the

author of an opinion has significant power to structure the announced rule. When the bench divides into two distinct factions, a dynamic similar to the one modeled may emerge. But extant models of opinion assignment are formulated in policy space rather than case space so the conclusions of our model do not follow immediately.

A Appendix

Proof of Lemma 1. We prove the more general result discussed in footnote 17. For concreteness, consider a judge from faction L. Let z^L be any threshold. Then:

$$v^L \left(\bar{y}^L, z^R \right) - v^L \left(z^L, z^R \right) = p \left| \int_{y^L}^{z^L} \eta \left(x; \bar{y}^L \right) dF \left(x \right) \right| \geq 0$$

which implies that implementing her ideal cut-point derives at least as much utility as any other cut-point, *ceteris paribus.* Furthermore, since $\eta \left(x; \bar{y}^L \right) > 0$ for all $x \neq \bar{y}^L$ and since F is weakly increasing, the expression must be strictly positive whenever $F \left(z^L \right) \neq F \left(y^L \right)$. Hence z^L is optimal only if $F \left(z^L \right) = F \left(\bar{y}^L \right)$. Moreover, if $\bar{y}^L \in int \left\{ supp \left(F \right) \right\}$, then F is strictly increasing in the neighborhood of $\bar{y}^L$, and so $F \left(z^L \right) = F \left(\bar{y}^L \right)$ implies $z^L = \bar{y}^L$. $\square$

Proof of Lemma 2. (Only if $\Rightarrow$). Suppose $F \left(z^L \right) < F \left(\bar{y}^L \right)$. Then, since $\eta \left(x; \bar{y}^i \right) > 0$ for all $x \neq \bar{y}^i$ and since F is weakly increasing:

$$v^L \left(\bar{y}^L, z^R \right) - v^L \left(z^L, z^R \right) = p \int_{z^L}^{y^L} \eta \left(x; \bar{y}^L \right) dF \left(x \right) > 0$$

and:

$$v^R \left(\bar{y}^L, z^R \right) - v^R \left(z^L, z^R \right) = (1 - p) \int_{z^L}^{y^R} \eta \left(x; \bar{y}^R \right) dF \left(x \right) > 0$$

which implies that $\left(\bar{y}^L, z^R \right)$ is a Pareto improvement over $\left(z^L, z^R \right)$. Hence, if $\left(z^L, z^R \right)$ is Pareto optimal, $F \left(z^L \right) \geq F \left(\bar{y}^L \right)$. Using similar arguments, we can show that $F \left(\bar{y}^L \right) \leq$

$F\left(z^i\right) \leq F\left(\bar{y}^R\right)$ for $i \in \{L, R\}$.

Now, suppose $F\left(z^L\right) \neq F\left(z^R\right)$. WLOG suppose $z^L < z^R$ and let $z^C \in \left(z^L, z^R\right)$ be such that $F\left(z^C\right) = pF\left(z^L\right) + (1-p)F\left(z^R\right)$. By the continuity of F, such a z^C must exist. Then:

$$
\begin{aligned}
v^L\left(z^C, z^C\right) - v^L\left(z^L, z^R\right) &= (1-p)\int_{z^C}^{z^R} \eta\left(x; \bar{y}^L\right) dF(x) - p\int_{z^L}^{z^C} \eta\left(x; \bar{y}^L\right) dF(x) \\
&> (1-p)\int_{z^C}^{z^R} \eta\left(z^C; \bar{y}^L\right) dF(x) - p\int_{z^L}^{z^C} \eta\left(z^C; \bar{y}^L\right) dF(x) \\
&= \eta\left(z^C; \bar{y}^L\right)\left[pF\left(z^L\right) + (1-p)F\left(z^R\right) - F\left(z^C\right)\right] \\
&= 0
\end{aligned}
$$

where the second line makes use of the fact that $\eta\left(\cdot; \bar{y}^L\right)$ is strictly increasing on the interval $\left[z^L, z^R\right]$, that $F\left(z^L\right) < F\left(z^C\right) < F\left(z^R\right)$, and that F is weakly increasing. By a symmetric argument, we can show that $v^R\left(z^C, z^C\right) > v^R\left(z^L, z^R\right)$. Hence $\left(z^C, z^C\right)$ is a Pareto improvement over $\left(z^L, z^R\right)$. The proof follows by contrapositive.

(If $\Leftarrow$). Let $F\left(\bar{y}^L\right) \leq F\left(z^L\right) = F\left(z^R\right) \leq F\left(\bar{y}^R\right)$, but suppose that $\left(z^L, z^R\right)$ is not Pareto optimal. Then there exists some other pair $\left(z^{L'}, z^{R'}\right)$ which is a Pareto improvement over $\left(z^L, z^R\right)$. Clearly, if $F\left(z^{L'}\right) = F\left(z^L\right)$ and $F\left(z^{R'}\right) = F\left(z^R\right)$, then $v^L\left(z^{L'}, z^{R'}\right) = v^L\left(z^L, z^R\right)$ and $v^R\left(z^{L'}, z^{R'}\right) = v^R\left(z^L, z^R\right)$, which is not a Pareto improvement. Hence, either $F\left(z^{L'}\right) \neq F\left(z^L\right)$ or $F\left(z^{R'}\right) \neq F\left(z^R\right)$.

Suppose $\left[F\left(z^{L'}\right) - F\left(z^L\right)\right]\left[F\left(z^{R'}\right) - F\left(z^R\right)\right] \geq 0$. Recall that v^L is decreasing in z^L and z^R (whenever these are above $\bar{y}^L$) and v^R is increasing in z^L and z^R (whenever these are below $\bar{y}^R$). It follows that either $v^L\left(z^{L'}, z^{R'}\right) < v^L\left(z^L, z^R\right)$ or $v^R\left(z^{L'}, z^{R'}\right) < v^R\left(z^L, z^R\right)$. Hence, we must have $\left[F\left(z^{L'}\right) - F\left(z^L\right)\right]\left[F\left(z^{R'}\right) - F\left(z^R\right)\right] < 0$. Since $F\left(z^L\right) = F\left(z^R\right)$, this implies $z^{L'} \neq z^{R'}$.

WLOG, suppose $z^{L'} < z^{R'}$. (A symmetric argument holds for the reverse case.) Let $\hat{z} \in \left(z^{L'}, z^{R'}\right)$ be s.t. $F(\hat{z}) = pF\left(z^{L'}\right) + (1-p)F\left(z^{R'}\right)$. If $F(\hat{z}) \geq F\left(z^L\right) = F\left(z^R\right)$, then we

know that $v^L(\hat{z}, \hat{z}) \leq v^L(z^L, z^R)$. Moreover, by arguments in an earlier part of this proof, we know that $v^L(\hat{z}, \hat{z}) > v^L(z^{L'}, z^{R'})$. Then, since $v^L(z^{L'}, z^{R'}) > v^L(z^L, z^R)$, it must be that $F(\hat{z}) < F(z^L)$. By a symmetric argument for R, we can conclude that $F(\hat{z}) > F(z^L)$, which is a contradiction. $\qquad\square$

Proof of Proposition 1. The first part of the proposition follows immediately from the discussion that precedes it. It remains to prove the enumerated points. First, we show that, as defined, $\delta(z^C) > 0$ for all z^C. It suffices to show that the denominators of both δ^L and δ^R are always positive. To see this for δ^L, note that:

$$\eta\left(z^C; \bar{y}^L\right) + v^L\left(z^C, z^C\right) - v^L\left(\bar{y}^L, \bar{y}^R\right)$$

$$= \eta\left(z^C; \bar{y}^L\right) - p\int_{\bar{y}^L}^{z^C} \eta\left(x; \bar{y}^L\right) dF(x) + (1-p)\int_{z^C}^{\bar{y}^R} \eta\left(x; \bar{y}^L\right) dF(x)$$

$$\geq \eta\left(z^C; \bar{y}^L\right) - p\int_{\bar{y}^L}^{z^C} \eta\left(x; \bar{y}^L\right) dF(x)$$

$$\geq \eta\left(z^C; \bar{y}^L\right) - p\left[F\left(z^C\right) - F(\bar{y})\right]\eta\left(z^C; \bar{y}^L\right)$$

$$> 0$$

where the 4th line makes use of the *IDID* property. A similar approach verifies the result for δ^R.

To show (1), note that $\delta(z^C) > 1$ provided that $\Delta v_i(z^C) = v^i(z^C, z^C) - v^i(\bar{y}^L, \bar{y}^R) < 0$ for either $i \in \{L, R\}$. Since η and F are both continuous, so are $v^i(z^C, z^C)$ are $\Delta v_i(z^C)$. Now, $\Delta v_L(z^C)$ is decreasing in z^C. (To see this, take $z' > z$. Then $\Delta v_L(z') = \Delta v_L(z) - \int_z^{z'} \eta(x; \bar{y}^i) dF(x) \leq \Delta v_L(z)$, since $\eta \geq 0$.) Furthermore, $\Delta v_L(\bar{y}^L) > 0$ and $\Delta v_L(\bar{y}^R) < 0$. Then, by the intermediate value theorem, there exists a unique $\bar{z} \in (\bar{y}^L, \bar{y}^R)$ s.t. $\Delta v_L(\bar{z}) = 0$, and $\Delta v_L(z) > 0$ provided that $z < \bar{z}$. Similarly, we can show that $\Delta v_R(z^C)$ is increasing in z^C, and that $\Delta v_R(\bar{y}^L) < 0$ and $\Delta v_R(\bar{y}^R) > 0$. Then, there exists a unique $\underline{z} \in (\bar{y}^L, \bar{y}^R)$ s.t. $\Delta v_R(\underline{z}) = 0$, and $\Delta v_R(z) > 0$ provided that $z > \underline{z}$.

It remains to show that $\underline{z} \leq \bar{z}$. Suppose not. Then for every $z^C \in \left[\bar{y}^L, \bar{y}^R\right]$, either $\Delta v_L\left(z^C\right) < 0$ or $\Delta v_R\left(z^C\right) < 0$. But this implies that there are no Pareto improvements over Autarky, which contradicts Lemma 2. Hence $\underline{z} \leq \bar{z}$.

To show (2), note that we have now shown that, for $i \in \{L, R\}$, $\Delta v_i\left(z^C\right) > 0$ for every $z^C \in (\underline{z}, \bar{z})$. Since:

$$\delta^i\left(z^C\right) = \frac{1}{1 + \frac{\Delta v_i(z^C)}{\eta(z^C; \bar{y}^i)}}$$

then $\delta^i\left(z^C\right)$ is increasing in z^C iff $\frac{\Delta v_i\left(z^C\right)}{\eta(z^C; y^i)}$ is decreasing in z^C. Now, we previously showed that Δv_L is decreasing and Δv_R is increasing in z^C, for $z^C \in \left[\bar{y}^L, \bar{y}^R\right]$. Moreover, in this region, $\eta\left(z^C; \bar{y}^L\right)$ is increasing and $\eta\left(z^C; \bar{y}^R\right)$ is decreasing. Hence $\frac{\Delta v_L\left(z^C\right)}{\eta(z^C; y^L)}$ is decreasing and so $\delta^L\left(z^C\right)$ is increasing in z^C. Likewise, $\frac{\Delta v_R\left(z^C\right)}{\eta(z^C; y^R)}$ is increasing and so $\delta^R\left(z^C\right)$ is decreasing in z^C.

Finally, let $D\left(z^C\right) = \delta^L\left(z^C\right) - \delta^R\left(z^C\right)$. We have shown that $D\left(z^C\right)$ is strictly increasing in z^C. By construction, $\delta^L\left(\bar{z}\right) = 1$ and $\delta^R\left(\underline{z}\right) = 1$. Since δ^L (resp. δ^R) is increasing (resp. decreasing), it follows that $\delta^L\left(\underline{z}\right) < \delta^R\left(\underline{z}\right) = 1$ and $\delta^R\left(\bar{z}\right) < \delta^L\left(\bar{z}\right) = 1$. Hence $D\left(\underline{z}\right) < 0$ and $D\left(\bar{z}\right) > 0$. Furthermore, since η and F are continuous, so are δ^i and D. Then, by the intermediate value theorem, there exists a unique $z^* \in (\underline{z}, \bar{z})$ s.t. $D\left(z^*\right) = 0$. Now, since $\delta\left(z^C\right) = \max\left\{\delta^L\left(z^C\right), \delta^R\left(z^C\right)\right\}$, the previous discussion implies:

$$\delta\left(z^C\right) = \begin{cases} \delta^R\left(z^C\right) & z^C < z^* \\ \delta^L\left(z^C\right) & z^C > z^* \end{cases}$$

. Then, by the properties of δ^L and δ^R, $\delta\left(z^C\right)$ is decreasing on $(\underline{z}, z^*)$ and increasing on $(z^*, \bar{z})$. Finally, this implies $\delta\left(z^C\right) \geq \delta\left(z^*\right) = \delta^*$ for all z^C. $\qquad \square$

Proof of Lemma 3. To see (1), recall from the proof of Proposition 1 that $\delta^i\left(z^C\right)$ is strictly decreasing in $\frac{\Delta v_i\left(z^C\right)}{\eta(z^C; \bar{y}^i)}$. Moreover, note that the denominator of this expression is independent

45

of p, and that:

$$\Delta v_i\left(z^C\right) = sgn\left(z^C - \bar{y}^i\right)\left[\int_{z^C}^{y^R} \eta\left(x;\bar{y}^i\right)dF\left(x\right) - p\int_{y^L}^{y^R} \eta\left(x;\bar{y}^i\right)dF\left(x\right)\right]$$

. Since $\int_{y^L}^{y^R} \eta\left(x;\bar{y}^i\right)dF\left(x\right) > 0$, then $\Delta v_L\left(z^C\right)$ is strictly decreasing in p and $\Delta v_R\left(z^C\right)$ is strictly increasing in p. Hence, $\delta^L\left(z^C\right)$ and $\delta^R\left(z^C\right)$ are strictly increasing and strictly decreasing in p, respectively.

Next, to see (2), recall that z^* is the solution to $D\left(z^C\right) = \delta^L\left(z^C\right) - \delta^R\left(z^C\right) = 0$. We showed in proposition 1 that δ^L and δ^R are increasing and decreasing in z^C, respectively, and so D is increasing in z^C. Moreover, part (1) of this lemma showed that D is strictly increasing in p. Now, let $p' > p$, and let $D\left(z^*;p\right) = 0$ and $D\left(z';p'\right) = 0$. Suppose $z' \geq z^*$. Then $0 = D\left(z';p'\right) \geq D\left(z^*;p'\right) > D\left(z^*;p\right) = 0$, which is a contradiction. Hence $z' < z^*$.

To see (4), note that $\underline{z}$ and $\bar{z}$ are defined by: $\delta^R\left(\underline{z}\right) = 1$ and $\delta^L\left(\bar{z}\right) = 1$. Let $p' > p$. Let $\underline{z} = \underline{z}\left(p\right)$ and $\underline{z}' = \underline{z}\left(p'\right)$. Similarly, let $\bar{z} = \bar{z}\left(p\right)$ and $\bar{z}' = \bar{z}\left(p'\right)$. Suppose $\underline{z}' \geq \underline{z}$. Then, since δ^R is increasing in p, $1 = \delta^R\left(\underline{z}',p'\right) \geq \delta^R\left(\underline{z};p'\right) > \delta^R\left(\underline{z};p\right) = 1$, which is a contradiction. Hence $\underline{z}' < \underline{z}$. An analogous argument shows that $\bar{z}' < \bar{z}$. $\qquad\square$

Proof of Lemma 4. To see (1), recall from the proof of Proposition 1 that $\delta^i\left(z^C\right)$ is strictly decreasing in $\frac{\Delta v_i\left(z^C\right)}{\eta\left(z^C;\bar{y}^i\right)}$. Again, note that the denominator of this expression is independent of ρ, and that the numerator can be re-written:

$$\rho\left\{sgn\left(z^C - \bar{y}^i\right)\left[\left(1-p\right)\int_{z^C}^{y^R} \eta\left(x;\bar{y}^i\right)dF\left(x;\left[\bar{y}^L,\bar{y}^R\right]\right) - p\int_{y^L}^{z^C} \eta\left(x;\bar{y}^i\right)dF\left(x;\left[\bar{y}^L,\bar{y}^R\right]\right)\right]\right\}$$

Now, since we are in the region of gains from trade, $\Delta v^i\left(z^C\right) \geq 0$, and so the term in braces must be non-negative. Further note that the term in braces only depends on the conditional distribution of F, and so is independent of ρ. Hence Δv_i is increasing in ρ, and so δ^i is weakly decreasing in ρ.

Part (3) and the results concerning $\bar{z}$ and $\underline{z}$ in part (2) follow straightforwardly, using similar arguments to those in the proof of Lemma 3. To show the result about z^* in part (2), note that z^* solves $\frac{\Delta v_L(z^C)}{\eta(z^C;\bar{y}^L)} = \frac{\Delta v_R(z^C)}{\eta(z^C;\bar{y}^R)}$. Let $\varphi_i(z^C)$ denote the term in braces, so that $\Delta v_i(z^C) = \rho\varphi_i(z^C)$. Then, z^* is the solution to:

$$\frac{\varphi_L(z^C)}{\eta(z^C;\bar{y}^L)} = \frac{\varphi_R(z^C)}{\eta(z^C;\bar{y}^R)}$$

. But since φ_i and η are both independent of ρ, it follows that z^* must be as well. $\qquad\square$

Proof of Proposition 2. Recall that (z^L, z^R) is sustainable at δ if $\phi^i(z^L, z^R, \delta) = -\eta(z^i; \bar{y}^i) + \frac{\delta}{1-\delta}\Delta v^i(z^L, z^R) \geq 0$ for $i \in \{L, R\}$. Since η is continuous and v^i are both continuous, then ϕ^i is jointly continuous in (z^L, z^R, δ). To show (1), suppose $(z^L, z^R) \in Z(\delta)$. Then $\phi^i(z^L, z^R, \delta) \geq 0$ which implies $\Delta v^i(z^L, z^R) \geq 0$ (since $\eta \geq 0$). Then, if $\delta' > \delta$, $\phi^i(z^L, z^R, \delta') \geq \phi^i(z^L, z^R, \delta) \geq 0$ (and the first inequality is strict if $\Delta v^i(z^L, z^R) > 0$). Hence $(z^L, z^R) \in Z(\delta')$, and so $Z(\delta) \subset Z(\delta')$.

To show (2), note that since $\eta(\bar{y}^i; \bar{y}^i) = 0$, $\phi^i(\bar{y}^L, \bar{y}^R, \delta) = 0$ for every $\delta \in [0, 1]$. Hence $(\bar{y}^L, \bar{y}^R) \in Z(\delta)$ for every δ.

To show (4), first note that by construction, $\phi^i(z^*, z^*, \delta^*) = 0$ for each i. Take some $\varepsilon > 0$, and let $z^{L'} = z^* - (1-p)\varepsilon$ and $z^{R'} = z^* + p\varepsilon$. Note that $z^{L'} \neq z^{R'}$. We seek to show that $\phi^i(z^{L'}, z^{R'}, \delta^*) > 0$. For concreteness, take $i = L$. Then:

$$\frac{d}{d\varepsilon}\phi^L(z^{L'}, z^{R'}, \delta^*) = (1-p)\eta'(z^{L'}; \bar{y}^L) + \frac{\delta^*}{1-\delta^*}p(1-p)\left[\eta(z^{L'}; \bar{y}^L)f(z^{L'}) - \eta(z^{R'}; \bar{y}^L)f(z^{R'})\right]$$

and so $\frac{d}{d\varepsilon}\phi^L(z^{L'}, z^{R'}, \delta^*)|_{\varepsilon=0} = (1-p)\eta'(z^*; \bar{y}^L) > 0$. Hence, for $\varepsilon > 0$ small enough, $\phi^L(z^{L'}, z^{R'}, \delta^*) > \phi^L(z^*, z^*, \delta^*) = 0$. We can similarly show that $\phi^R(z^{L'}, z^{R'}, \delta^*) > 0$. Now, since ϕ^i is continuous and increasing in δ, there exists $\delta' < \delta^*$, such that $(z^{L'}, z^{R'}) \in Z(\delta')$. But by construction, we know that there is no complete stare decisis regime (z', z') for which $(z', z') \in Z(\delta')$. This proves (4).

Finally, to show (3), let $\delta^{**} = \sup\left\{\delta | Z(\delta) = \left\{\left(\bar{y}^L, \bar{y}^R\right)\right\}\right\}$. Since $Z(0) = \left\{\left(\bar{y}^L, \bar{y}^R\right)\right\}$, the sup is well defined. Moreover, $\delta^{**} \in [0, \delta^*)$, since $\delta^{**} \leq \delta' < \delta^*$. We now show the relationship between δ^{**} and $\eta'\left(\bar{y}^i; \bar{y}^i\right)$. First, note that $\phi^i\left(\bar{y}^L, \bar{y}^R, \delta\right) = 0$. Now, define $z^{L''} = \bar{y}^L + (1-\pi)\varepsilon'$ and $z^{R''} = \bar{y}^R - \pi\varepsilon'$, for some $\pi \in [0, 1]$ and $\varepsilon' > 0$. By choosing π appropriately, $\left(z^{L''}, z^{R''}\right)$ spans every rule pair in an $\varepsilon'-$neighborhood of $\left(\bar{y}^L, \bar{y}^R\right)$. Again we focus on ϕ^L, and note that the proof for ϕ^R is analogous. Then:

$$\frac{d}{d\varepsilon'}\phi^L\left(z^{L''}, z^{R''}, \delta\right) = (1-\pi)\eta'\left(z^{L''}; \bar{y}^L\right) + \frac{\delta}{1-\delta}\left[(1-\pi)p\eta\left(z^{L''}; \bar{y}^L\right) - \pi(1-p)\eta\left(z^{R''}; \bar{y}^L\right)\right]$$

and so $\frac{d}{d\varepsilon'}\phi^L\left(z^{L''}, z^{R''}, \delta\right)|_{\varepsilon'=0} = -(1-\pi)\eta'\left(\bar{y}^L; \bar{y}^L\right) + \frac{\delta}{1-\delta}\pi(1-p)\eta\left(\bar{y}^R; \bar{y}^L\right)$. Now, if $\eta' = 0$, then $\frac{d}{d\varepsilon'} > 0$, and so $\left(z^{L''}, z^{R''}\right) \in Z(\delta)$ whenever $\left(z^{L''}, z^{R''}\right)$ is in an ε'-neighborhood of $\left(\bar{y}^L, \bar{y}^R\right)$. We have $\delta^{**} = 0$. Suppose instead $\eta' > 0$. Then, for δ small enough $\frac{d}{d\varepsilon'} < 0$, and so for every $\left(z^{L''}, z^{R''}\right)$ in the neighborhood of $\left(\bar{y}^L, \bar{y}^R\right)$, $\left(z^{L''}, z^{R''}\right) \notin Z(\delta)$. Hence $\delta^{**} > 0$. □

Proof of Proposition 3. Suppose judge i is the agenda setter. Her problem is: $\max_{(z^L, z^R)} v^i\left(z^L, z^R\right)$ s.t. $\left(z^L, z^R\right) \in Z(\delta)$. By Proposition 2, we know that $\left(z^L, z^R\right) \in Z(\delta)$ provided that $\phi^i\left(z^L, z^R, \delta\right) \geq 0$ for each $i \in \{L, R\}$. and $z^R \geq z^R\left(z^L; \phi^R = 0\right)$. Suppose $\delta > \delta^{**}$ (otherwise, autarky is the only sustainable equilibrium, and so that is what is chosen.) Intuitively, the agenda setter holds judges of the other faction to indifference. Hence, the problem becomes: $\max_{(z^L, z^R)} v^i\left(z^L, z^R\right)$ s.t. $\phi^{-i}\left(z^L, z^R, \delta\right) \geq 0$.

Forming the Lagrangian, and solving for the first order conditions implies: $\frac{v_L^i}{v_R^i} = \frac{\partial\phi^{-i}/\partial z^L}{\partial\phi^{-i}/\partial z^L}$, where $v_j^i = \frac{\partial v^i}{\partial z^j}$. Recall that: $\frac{v_L^i}{v_R^i} = \frac{p}{(1-p)}\frac{\eta\left(z^L; \bar{y}^i\right)f\left(z^L\right)}{\eta\left(z^R; \bar{y}^i\right)f\left(z^R\right)}$. Moreover, if $i = R$, then

$$\frac{\partial\phi^{-i}/\partial z^L}{\partial\phi^{-i}/\partial z^R} = \frac{\frac{1-\delta}{\delta}\eta'\left(z^L; \bar{y}^L\right) + p\eta\left(z^L; \bar{y}^L\right)f\left(z^L\right)}{(1-p)\eta\left(z^R; \bar{y}^L\right)f\left(z^R\right)}$$

and if $i = L$, then:

$$\frac{\partial\phi^{-i}/\partial z^L}{\partial\phi^{-i}/\partial z^R} = \frac{p\eta\left(z^L; \bar{y}^R\right)f\left(z^L\right)}{\frac{1-\delta}{\delta}\eta'\left(z^L; \bar{y}^L\right) + (1-p)\eta\left(z^R; \bar{y}^R\right)f\left(z^R\right)}$$

. Let $\left(\hat{z}^L, \hat{z}^R\right)$ be a maximizer. If $\hat{z}^L = \hat{z}^R$ (i.e. if the optimum is characterized by a complete stare decisis regime), then $\frac{v_L^i}{v_R^i} = \frac{p}{1-p}$. But, by inspection, it is clear that $\frac{\partial \phi^{-i}/\partial z^L}{\partial \phi^{-i}/\partial z^R}|_{z^L=z^R} = \frac{p}{1-p}$ iff $\delta = 1$. Hence, a complete stare decisis regime is chosen only if $\delta = 1$. $\qquad \square$

References

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

Baron, D. P. and Ferejohn, J. A. (1989). Bargaining in legislatures. *American Political Science Review*, 83(04):1181–1206.

Blume, L. E. and Rubinfeld, D. L. (1982). The dynamics of the legal process. *The Journal of Legal Studies*, 11(2):405–419.

Brenner, S. and Stier, M. (1996). Retesting segal and spaeth's stare decisis model. *American Journal of Political Science*, 40(4):1036–1048.

Brisbin, R. A. (1996). Slaying the dragon: Segal, spaeth and the function of law in supreme court decision making. *American Journal of Political Science*, 40(4):1004–1017.

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

Cameron, C. M. and Korhnauser, L. A. (2017). Chapter 3: What do judges want? how to model judicial preferences. *NYU Law and Economics Research Paper*.

Cross, R. and Harris, J. W. (1991). *Precedent in English Law*. Clarendon Press.

Duxbury, N. (2008). *The Nature and Authority of Precedent*. Cambridge University Press.

Fischman, J. B. (2011). Estimating preferences of circuit judges: A model of consensus voting. *The Journal of Law and Economics*, 54(4):781–809.

Gely, R. (1998). Of sinking and escalating: A somewhat new look at stare decisis. *University of Pittsburgh Law Review*, 60:89.

Goodhart, A. L. (1930). Determining the ratio decidendi of a case. *The Yale Law Journal*, 40(2):161–183.

Heiner, R. A. (1986). Imperfect decisions and the law: On the evolution of legal precedent and rules. *The Journal of Legal Studies*, 15(2):227–261.

Jovanovic, B. (1988). Rules vs. discretion in the legal process. *mimeo*.

Knight, J. and Epstein, L. (1996). The norm of stare decisis. *American Journal of Political Science*, pages 1018–1035.

Kornhauser, L. A. (1989). An ecomomic perspective on stare decisis. *Chicago-Kent Law Review*, 65:63.

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

Kornhauser, L. A. (1998). Stare decisis. *The New Palgrave Dictionary of Economics and the Law*, 3:509–14.

Lax, J. R. and Cameron, C. M. (2007). Bargaining and opinion assignment on the us supreme court. *Journal of Law, Economics, and Organization*, 23(2):276–302.

Levi, E. H. (2013). *An Introduction to Legal Reasoning*. University of Chicago Press.

Llewellyn, K. N. (2012). *The Bramble Bush: On Our Law and its Study*. Quid Pro Books.

MacCormick, N. and Summers, R. S. (1997). Further general reflections and conclusions. *Interpreting Precedents: A Comparative Study. Aldershot [etc.]: Ashgate/Dartmouth*.

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

Postema, G. J. (1986). *Bentham and the Common Law Tradition.*

Rasmusen, E. (1994). Judicial legitimacy as a repeated game. *Journal of Law, Economics, and Organization*, pages 63–83.

Romer, T. and Rosenthal, H. (1978). Political resource allocation, controlled agendas, and the status quo. *Public choice*, 33(4):27–43.

Schauer, F. (1987). Precedent. *Stanford Law Review*, pages 571–605.

Songer, D. R. and Lindquist, S. A. (1996). Not the whole story: The impact of justices' values on supreme court decision making. *American Journal of Political Science*, 40(4):1049–1063.

Spaeth, H. J. and Segal, J. A. (2001). *Majority Rule or Minority Will: Adherence to Precedent on the US Supreme Court.* Cambridge University Press.

Wasserstrom, S. J. and Seidman, L. M. (1988). The fourth amendment as constitutional theory. *Georgetown Law Journal*, 77:19.