UNIVERSITÀ CATTOLICA del Sacro Cuore

Dipartimento di Economia Internazionale, delle Istituzioni e dello Sviluppo (DISEIS)

WORKING PAPER SERIES

WP 2602

March 2026

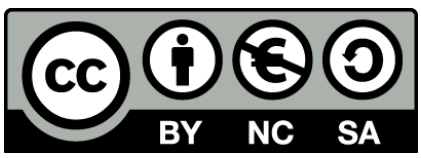
Reversal Costs and Executive Overreach

Barbara Antonioli Mantegazzini, Federico Trombetta

Prior to being published, DISEIS working papers are subject to an anonymous refereeing process by members of the editorial board or by members of the department.

DISEIS Working Papers often represent preliminary work which are circulated for discussion and comment purposes. Citation and use of such a paper should account for its provisional character. A revised version may be available directly from the author.

DISEIS WP 2602



DISEIS
Dipartimento di Economia Internazionale, delle Istituzioni e dello Sviluppo
Università Cattolica del Sacro Cuore
Via Necchi 5
20123
Milano

Reversal Costs and Executive Overreach*

Barbara Antonioli Mantegazzini† & Federico Trombetta‡

March 9, 2026

Abstract

Executives may implement legally contestable policies aggressively before courts reach a final legality determination, creating reliance and other sunk effects that make reversal costly. We study a complete-information sequential game in which an executive chooses policy aggressiveness and a court then decides whether to uphold the policy or strike it. If reversal costs increase sufficiently convexly in aggressiveness, the court strikes mild policies but upholds sufficiently aggressive ones to avoid disruption. Anticipating this, the executive overreaches—choosing a policy more extreme than its ideal point—to deter full reversal, yielding inefficient excess implementation relative to a commitment benchmark. Institutions that limit pre-review sunk effects (stays, phased implementation, expedited review) mitigate this distortion.

JEL codes: D72; D74; K23; K40; P16.

Keywords: judicial review; reversal costs; executive overreach; strategic entrenchment; tariffs.

*We thank Daniele Senzani, Georg Vanberg and Marco Zaffalon for their insightful comments and suggestions.

†Full Professor in Public Economics and in Sustainability Policies; Head of the Competence Centre on Labour, Welfare and Social Research; Department of Business Economics, Health and Social Care, University of Applied Sciences and Arts of Southern Switzerland, Manno. Email: barbara.antonioli@supsi.ch.

‡Assistant Professor in Economics; Department of International Economics, Institutions and Development, Universitá Cattolica del Sacro Cuore, Milano. Email: federico.trombetta@unicatt.it.

"If you win, tell me how the reimbursement process would work. Would it be a complete mess? ... It seems to me like it could be a mess."
(Justice Amy Coney Barrett)[1]

1 Introduction

Since Hamilton's discussion of the judiciary as a moderating "check,"[2] judicial review is often viewed as a checks-and-balances device that constrains—and thereby tends to moderate—the actions of the political branches, preserving freedom (La Porta et al., 2004). We show that, once reversal costs are taken into account, judicial review can instead distort policy *toward* more aggressive implementation rather than disciplining it.

In particular, executives may implement legally contestable policies in ways that make judicial reversal costly. A salient example is tariffs. An administration can announce and enforce tariffs quickly, triggering immediate price adjustments, supply-chain reallocation, compliance investments, and lobbying by newly protected industries. Even if litigation is filed promptly, courts may confront a remedial dilemma once implementation has begun: unwinding tariffs can create disruption, refund obligations, and administrability problems that are absent ex ante. As a result, judicial control may operate not only through legality determinations but also through remedial choices that implicitly weigh the costs of reversal.

This paper studies how *reversal costs* reshape executive incentives under judicial review. The central idea is simple: when courts cannot commit ex ante to fully unwind an unlawful policy after reliance has formed, the executive can strategically increase reliance—by choosing a more aggressive, faster, or broader implementation—knowing that once implementation generates sufficient disruption the court's remedial calculus may favor leaving the policy in place. We formalize this mechanism in a two-player sequential game between an Executive and a Court. The Executive chooses policy aggressiveness, a reduced-form index capturing legal risk/interpretive stretch, implementation scale and speed, and the extent to which early implementation generates entrenchment. After observing aggressiveness, the Court chooses whether to *uphold* (allow the policy to remain in effect, including via limited or prospective remedies) or *strike* (fully unwind). Upholding generates a legality/precedent loss that increases in aggressiveness; striking generates a reversal/disruption loss that also increases in aggressiveness. Importantly, the model does not assume that courts reinterpret legality in light of disruption. Rather,

[1] Justice Amy Coney Barrett addressing Neal Katyal, a lawyer for the challengers against tariffs. Tr. of Oral Arg. at 153–55, *Learning Resources, Inc. v. Trump*, No. 24-1287 (U.S. Nov. 5, 2025), `https://www.supremecourt.gov/oral_arguments/argument_transcripts/2025/24-1287_b07d.pdf`.

[2] Hamilton, *Federalist No. 78*, in Hamilton (1961): "It ... serves to moderate the immediate mischiefs ... [and] operates as a check upon the legislative body ...".

legality concerns are captured by the court's legality loss, while reversal costs arise at the remedial stage once implementation has created reliance. The court's disposition therefore reflects the institutional consequences of alternative remedies rather than a departure from legal reasoning.

The model delivers two sharp results. First, when reversal costs rise *more convexly* than the court's legality loss, the Court exhibits an "entrenchment" decision rule: it is willing to strike relatively mild policies but prefers to uphold sufficiently aggressive ones because disruption from unwinding becomes too large. Anticipating this, the Executive may choose a policy *more extreme than its preferred one* in order to cross the Court's upholding threshold. We call this equilibrium behavior *executive overreach*. Overreach arises for an intermediate range of preference divergence between the Court and the Executive: when divergence is small the Executive refrains from acting to avoid a strike, and when divergence is large the preferred policy is sustained or judicial review is effectively nonbinding, but for intermediate divergence the Executive optimally "buys" judicial acquiescence by increasing aggressiveness and thereby raising reversal costs. Second, when reversal costs are relatively less convex than the legality loss, judicial review induces *moderation* instead: the Executive either implements its preferred policy when divergence is small or scales back to the largest policy the Court is willing to uphold when divergence is large. In this regime, overreach does not occur.

These strategic effects overturn standard intuitions about the welfare consequences of judicial review. In the overreach regime, equilibrium outcomes are inefficient because the Executive distorts policy away from its preferred interior level solely to manipulate the Court's ex post remedial incentives. Both actors would strictly prefer the moderate implement-and-uphold outcome that would obtain if the Court could credibly commit to uphold moderate policies. From the perspective of a representative voter whose policy objective aligns with the Court's (e.g. a voter harmed by protectionist tariffs), the welfare effect of judicial review is non-monotone in executive bias: review is beneficial when it deters contested policy, irrelevant when it does not bind, and harmful precisely in the intermediate region where it induces overreach.

Related work Our analysis speaks to the positive political economy of courts and separation of powers, where judicial decisions are strategic choices under political and institutional constraints. Classic rational-choice accounts emphasize that courts anticipate reactions by elected actors and tailor rulings accordingly (Gely and Spiller, 1990; Spiller and Gely, 1992). A related model of court-legislatures strategic interactions stresses the role of the surrounding political environment, and the weight courts place on not being defied by legislatures (Vanberg, 2001). Legislatures may also use judicial review strategically, in order to shape and improve policy outcomes (Shipan, 2000). A prominent formal mechanism is that judicial review can discipline (or sometimes amplify) electoral "pos-

turing" by elected officials (Fox and Stephenson, 2011), and it can have nontrivial welfare effects depending on whom review is designed to protect, sometimes incentivizing more extreme policy proposal for signaling reasons (Fox and Stephenson, 2015). Furthermore, the prospect of reversal may induce governments to invest ex ante in justification and record-building under hard-look review, for signaling purposes, overcoming the court's informational disadvantage (Stephenson, 2006).[3] Relative to these contributions, our mechanism is distinct: the executive's strategic variable is the deliberate creation of *sunk effects* that change the court's *remedial* incentives ex post, in a game of complete information.

The model also relates to formal theories of *ex post oversight* and judicial control of agencies. In Bueno de Mesquita and Stephenson (2007) and subsequent work, oversight affects how agencies allocate effort across observable and unobservable dimensions, potentially distorting incentives even when oversight improves expected policy quality. Likewise, Turner (2017), Turner (2019), and Patty and Turner (2021) study how the prospect of review (often under informational frictions) shapes ex ante policy choice and justificatory or precision effort. Related work emphasizes how the timing and availability of review affect agencies' incentives (Seidenfeld, 1997). We share the ex post timing—review occurs after an action is taken—but we highlight a different distortion that operates even under complete information: when the *court* bears reversal costs, the executive may choose *more extreme* implementation precisely to raise the cost of unwinding and thereby induce judicial acquiescence. In terms of welfare, this implies that judicial review may be useful when preferences are fairly (but not completely) aligned, and irrelevant when they are very misaligned.

Methodologically, the paper also differs from models that locate strategic adaptation along other margins. In Spiller and Tiller (1997) and Tiller and Spiller (1999), agencies and courts interact strategically through the design of administrative process and the manipulation of *decision instruments* (e.g., procedural formality, rulemaking versus adjudication) that affect exposure to review; by contrast, our strategic variable is the *degree of entrenchment* created by implementation itself.

Finally, our results complement broader accounts of checks and balances that emphasize nontrivial tradeoffs in institutional design (Gratton and Morelli, 2022) and political-economy work in which repeal is endogenous and may be costly or strategically constrained (Delgado-Vega et al., 2025). Our contribution is to show how an analogous "cost of undoing" channel operates *within* judicial review: costly reversal makes review effectively state-contingent, and that state dependence can itself induce overreach. The institutional implication is that reducing pre-review sunk effects—through stays, phased implementation, escrow/refund mechanisms in tariff administration, or expedited

[3]Other contributions study courts' own choice of legal rules and doctrine under uncertainty and complexity (Clark, 2016; Callander and Clark, 2017).

review—can mitigate strategic entrenchment without requiring courts to ignore disruption at the remedial stage.

The remainder of the paper proceeds as follows. Section 2 presents the model. Section 3 characterizes the subgame perfect equilibria and isolates the parameter conditions under which overreach versus moderation arises. Section 4 studies welfare implications and the representative-voter perspective. Section 5 discusses the institutional implications and Section 6 concludes.

2 Model

We study a two-player game. The players are an Executive (E) and a Court (C). Timing is as follows: (i) E chooses an aggressiveness level of the policy $a \geq 0$; (ii) If $a = 0$, payoffs are paid and the game ends. If $a > 0$, C observes it and chooses a disposition $y \in \{Uphold(U), Strike(S)\}$. If $y = S$, the implemented policy reverts back to $a = 0$. Otherwise, it remains the policy proposed by E. (iii) payoffs are realized and the game ends. Our solution concept is pure-strategy Subgame Perfect Nash Equilibrium (SPNE).

Aggressiveness a is a reduced-form index capturing legal risk/interpretive stretch, implementation speed and scale, and the degree to which early implementation generates reliance and entrenchment. The Court faces a legality/precedent loss from sustaining the measure, $L(a)$, and a reversal/disruption loss from unwinding it after implementation has begun, $D(a)$.

We use "Uphold" as shorthand for judicial dispositions that allow the policy to remain in effect (including limited or prospective remedies), whereas "Strike" denotes dispositions that fully unwind implementation.

This single Court move is a reduced-form representation of multi-stage judicial control (e.g., interim relief, stays, final legality determinations, and remedy design): the key object is that the court's ultimate disposition reflects a tradeoff between legality/precedent concerns and the institutional disruption of reversal once reliance has formed.

2.1 Payoffs

We assume that both players have a symmetric, power-form and convex loss function that depends on the distance between the implemented policy and their favorite one. In Appendix B we show that the main result of this paper can be obtained with more general functional form assumptions. Here we assume the resulting (concave) payoff function is

$$\pi_I(a) = -\beta(|a - b_I|)^r$$

with $I \in \{C, E\}$ and $r > 1$.

We assume that the court's bliss point is $b_C = 0$ (therefore, the Court would always strike down the policy, absent additional considerations. This allows us to have $L(a) = \beta a^r$), while the executive prefers an interior level of the policy, i.e. $b_E = b > 0$. Therefore, b measures the bias, in favor of the policy, of the executive vis-à-vis the court. In case of tariffs, this implies assuming that the Court shares the objective function of the average voter, while the executive is biased in favor of a certain level of tariffs due to partisan considerations, influence by special interest groups and so on.

Note that, absent additional strategic considerations, E would always choose $a = b$. Therefore, the difference between the equilibrium a and b measures whether there is executive overreach ($a > b$) or underreach ($a < b$).

The main innovation of this model is the introduction of reversal costs. They are paid by C in case of $y = S$, and are increasing and convex in a. For illustrative purposes, we assume

$$D(a) = \gamma a^p$$

with $p > 1$. Intuitively, striking generates a reversal/disruption loss $D(a)$, increasing in pre-review sunk effects (reliance, administrability, and adjustment costs of unwinding implementation). They summarize legally relevant institutional losses that can arise under alternative dispositions and remedies even when legality is contested. Remedial doctrine routinely requires courts to consider administrability, reliance, and the feasibility of unwinding when choosing the form and scope of relief. These objects are reduced-form measures of institutional and social losses associated with each disposition, not private "tastes" for cost minimization.

2.2 Discussion of the assumptions

The model deliberately abstracts from many institutional details in order to isolate a single strategic force: when courts cannot commit to fully unwind an implemented policy, the executive can choose implementation in ways that make reversal increasingly unattractive. We capture this through a one-dimensional aggressiveness index a that bundles legal risk (interpretive stretch), speed and scale of roll-out, and the degree of early entrenchment. This reduced-form approach is useful because these margins typically move together in practice: expansive interpretations are often paired with rapid implementation, and rapid implementation accelerates reliance. The power-form payoff function $\pi_I(a) = -\beta|a - b_I|^r$ with $r > 1$ imposes standard concavity, ensuring a unique interior ideal point and making deviations increasingly costly; it is not essential for the qualitative results but yields transparent thresholds and comparative statics. Setting the Court's bliss point to zero provides a tractable benchmark in which, absent reversal costs, the Court prefers the no-policy status quo to any positive policy. Similar "overreach"

forces arise when $b_E > b_C$, which generates a potential conflict of objectives. However, with $b_C > 0$, the court may uphold moderate policies even without large reversal costs. Finally, representing judicial control as a single disposition $y \in \{U, S\}$ is a parsimonious way to summarize a sequence of procedural choices (stays, preliminary relief, merits, and remedy design) into the ultimate policy consequence: whether the implemented policy remains in effect or is fully unwound.

Two related substantive assumptions concern the reversal costs $D(a) = \gamma a^p$. First, one might object that the Court should focus solely on legality. In reality, however, judges often care directly about the consequences of their decisions, for at least two reasons. First, remedial and interim-relief doctrine gives courts discretion to account for the practical costs of unwinding unlawful action. The Supreme Court has emphasized that equitable relief is not automatic and that courts must consider disruption to the public and other parties: in *Weinberger v. Romero-Barcelo*, the Court explained that equity courts "should pay particular regard for the public consequences" of an injunction and reiterated that where an injunction would adversely affect the public interest, a court may withhold relief pending a final determination.[4] Consistently, in *eBay Inc. v. MercExchange, L.L.C.*, the Court reaffirmed that injunctive relief requires, inter alia, balancing hardships and ensuring "that the public interest would not be disserved" by an injunction.[5] In election litigation, the Court has treated late-stage judicial intervention itself as potentially disruptive, warning that court orders affecting elections can result in "confusion and consequent incentive to remain away from the polls," with the risk increasing as the election nears.[6] These administrability concerns also surface explicitly in the tariff context: at oral argument, Justice Barrett pressed challengers on the feasibility of post-invalidation refunds—asking whether the reimbursement process would be "a complete mess," and later remarking that "it could be a mess" and "So a mess?".[7] In dissent, Justice Kavanaugh likewise emphasized the practical stakes, noting that the United States may be required to refund billions and that "the refund process is likely to be a 'mess' ".[8] On top of this, it has to be noted that - sometimes - courts adopt explicit rules that reduce the problem of post-strike disruptions. For example, the German Federal Constitutional Court introduced the notion of a law "incompatible with the Basic Law" (i.e., the law has to be changed, but it remains in place until the change happens), as an alternative of declaring a law "null and void". France used to have only "abstract constitutional review", where Constitutional Court decisions were taken

[4] *Weinberger v. Romero-Barcelo*, 456 U.S. 305, 312–13 (1982).

[5] *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 391 (2006).

[6] *Purcell v. Gonzalez*, 549 U.S. 1, 4–5 (2006) (per curiam).

[7] Tr. of Oral Arg. 153–155, *Learning Resources, Inc. v. Trump*, No. 24-1287 (U.S. Nov. 5, 2025), available at `https://www.supremecourt.gov/oral_arguments/argument_transcripts/2025/24-1287_b07d.pdf`.

[8] *Learning Resources, Inc. v. Trump*, 607 U.S. ___ (2026) (Kavanaugh, J., dissenting) (slip op. at 6) (citing Tr. of Oral Arg. 153–155).

before the promulgation of the law (Vanberg, 2005). Second, as pointed out by Vanberg (2001), court decisions are rarely self-enforcing, and legislative bodies may try to defy those decisions. Courts, therefore, behave strategically, taking those additional actions into account. Striking down a policy with very high reversal costs is likely to be very unpopular, hence increasing the risk of defiance (and, therefore, making the courts more reluctant to act).

A second question that may arise is about the empirical plausibility of $r < p$, i.e. of the fact that reversal costs are *more convex* than the court's legality/precedent loss from upholding. There are several reasons to expect strong convexity in practice. First, reliance and adjustment often exhibit threshold effects: modest implementation may generate limited sunk investments, but once a policy reaches scale it triggers large fixed costs (compliance systems, reconfigured supply chains, procurement contracts, and organizational routines) that are costly to reverse. Second, disruption from unwinding can increase superlinearly because reversal is not merely the negative of implementation: undoing may require refunds, transition rules, adjudication of accrued rights, and coordination across agencies and private actors, with congestion and administrative capacity constraints that become binding at higher scale. Third, political and institutional frictions can magnify convexity. As entrenchment grows, the set of affected constituencies expands and becomes more organized, raising the political costs of reversal and increasing the probability that unwinding must be partial, delayed, or procedurally complex. These mechanisms are especially salient in trade policy. Small tariff changes may be absorbed with limited reallocation, while large, rapidly imposed tariffs can induce extensive re-sourcing, contract renegotiation, and inventory and pricing changes; unwinding then entails nontrivial administrability (customs reclassification, drawback and refund procedures) and disruption in precisely the domains that remedial doctrine instructs courts to consider. These mechanisms imply that marginal unwinding costs plausibly rise faster than marginal legality/precedent costs at scale, making $p > r$ a reasonable benchmark. Taken together, these considerations justify modeling $D(\cdot)$ as highly convex and motivate why the region $r < p$—the region that generates the possibility of executive overreach in equilibrium—may be a relevant empirical benchmark.

3 Equilibrium

We solve the game by backward induction.

3.1 Court decision rule

Given a, the Court applies a remedial decision rule that compares the institutional–legal loss from upholding to the institutional–social loss from reversal. This reduced-form rule

captures the remedial tradeoff courts face when implementation has generated reliance: the disposition (and especially the remedy) may be shaped by the costs of unwinding, without implying that legality is irrelevant. It upholds whenever $\pi_C(a, U) = \pi_C(a) \geq \pi_C(a, S) = \pi_C(0) - D(a)$. To guarantee pure-strategy existence of a SPNE for all parameter values, we assume that the Court breaks ties in favor of U.[9] Under the functional forms we assume, the choice depends on the comparison between r and p, as discussed in Lemma 1 below (all the proofs are in Appendix A). Throughout the paper, we assume $p \neq r$. The knife-edge case of $p = r$ is omitted for brevity.

Lemma 1 *In every SPNE where the Court breaks ties in favor of U:*

- *if $r < p$ and $a > 0$, the Court chooses $y = U$ iff $a \geq \hat{a}_1 := \left(\frac{\beta}{\gamma}\right)^{\frac{1}{p-r}}$;*
- *if $r > p$ and $a > 0$, the Court chooses $y = U$ iff $a \leq \hat{a}_2 := \left(\frac{\gamma}{\beta}\right)^{\frac{1}{r-p}}$;*

Lemma 1 contains an important insight for the rest of the paper. In the presence of reversal costs, the Court's choice crucially depends on the relative "shape" of those costs, compared with the loss of upholding a disliked policy. If $r < p$, meaning that the (negative of the) reversal cost function is flat at the beginning, and then steeper than $\pi_C(a)$, then C has an incentive to strike "mild" policies and to uphold those that are more extreme, given the steep increase of $D(a)$. Figure 1 illustrates this case for $r = 2$ and $p = 4$.

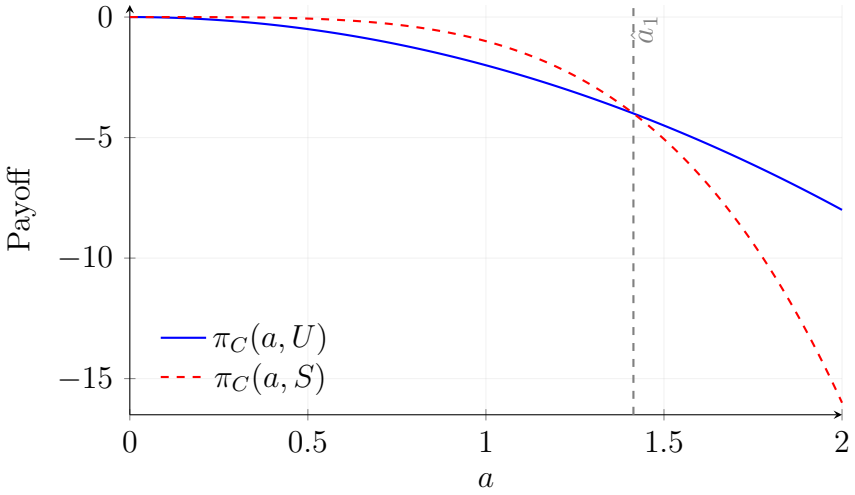


Figure 1: Payoffs as a function of a. Other parameters: $r = 2$, $p = 4$, $\beta = 2$, $\gamma = 1$.

One final point to note is that, although the functional form assumptions we are making greatly simplify the exposition and allow us to find a closed form solution, they are not necessary for this result. As long as, for sufficiently big a, $D(a)$ is (and remains)

[9] If the Court instead broke ties in favor of S, pure-strategy SPNEs would fail to exist for some parameter values.

steeper than the loss from the implementation of the policy, and it crosses from below at least once, then there will be a threshold in a above which the Court chooses $y = U$.

3.2 Executive choice

Anticipating the court's threshold $\hat{a}$, the executive chooses $a \geq 0$. Consider first the case of $p > r$. If $a \geq \hat{a}_1$, the policy is upheld and the executive obtains

$$\pi_E(a) = -\beta(|a - b|)^r$$

if instead $a < \hat{a}_1$ and therefore $y = S$, E obtains

$$\pi_E(0) = -\beta(|-b|)^r$$

As tie breaking rules, we assume that the executive pays an arbitrarily small "reputation cost" if $y = S$, and therefore chooses $a = 0$ if she anticipates $y = S$. Moreover, we assume that, when indifferent, E prefers the implementation of a higher level of a.[10]

Recalling that, absent additional constraints, E would always choose $a = b$, it is clear that the solution of E's decision problem depends on b. In particular, if $b \geq \hat{a}_1$, the preferred policy of the executive is upheld by the court, and therefore $a^* = b$. Consider now the opposite case. Choosing $a = b < \hat{a}$ is dominated by $a = 0$, because it would lead to $y = S$. Therefore, E can either opt for $a = 0$, or choose the smallest policy that leads to $y = U$, namely $a = \hat{a}_1$. Given the symmetry around b of E's payoff function, the choice depends on whether b is closer to $\hat{a}_1$ or to 0. Proposition 1 illustrates the strategies.

Proposition 1 *Assume $p > r$ and the tie-breaking rules stated above. In the unique SPNE of the game, E's strategy is as follows:*

$$a^* = \begin{cases} 0, & b < \frac{\hat{a}_1}{2}, \\ \hat{a}_1, & b \in \left[\frac{\hat{a}_1}{2}, \hat{a}_1\right], \\ b, & b > \hat{a}_1. \end{cases}$$

Figure 2 illustrates the executive strategy as a function of the bias. It is quite clear that, when $p > r$, we may observe executive overreach, i.e. a more extreme policy than the one E would like to obtain (and that would be obtained absent judicial review). This happens, in particular, for an intermediate bias. When the bias is small, instead, E prefers to avoid a court's strike, and therefore avoid the contested policy at all. In terms of comparative statics, therefore, executive overreach (measured as $a^* - b$) is non-monotonic in b, hence in the divergence of interests between the executive and the court.

[10] Those tie breaking rules are relevant only for uniqueness of equilibria in few knife-edge cases. Dropping them would just give rise to multiplicity in those cases.

Moreover, the range of b for which we observe overreach is increasing in $\hat{a}_1$, and therefore in the relative weight of the policy loss vis-a-vis the reversal cost.

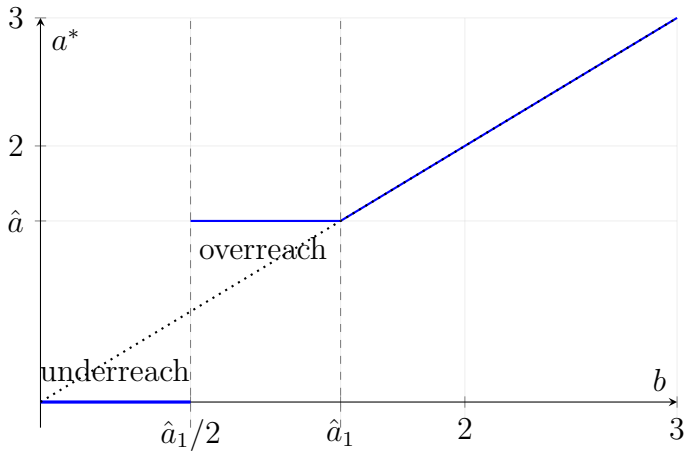


Figure 2: E's equilibrium strategy as a function of b. Other parameters: $r = 2$, $p = 4$, $\beta = 2$, $\gamma = 1$.

Proposition 2 shows that the relative shape of the reversal cost function and the policy loss function is crucial in determining whether overreach is observed in equilibrium.

Proposition 2 *Assume $p < r$. In the unique SPNE of the game, E's strategy is as follows:*

$$a^* = \begin{cases} b, & b < \hat{a}_2, \\ \hat{a}_2, & b \geq \hat{a}_2. \end{cases}$$

When $p < r$, executive overreach is not observed. E either chooses the preferred policy, for small biases, or moderates choosing $\hat{a}_2$, for larger b. Therefore, the Court is not able to avoid the policy altogether, but is able to induce moderation when the bias is large.

4 Welfare

In terms of welfare, we can show that, when there is executive overreach, the equilibrium is Pareto-dominated by the outcome $a = b$, $y = U$. Consider the court. In presence of overreach, C obtains a payoff of $-\beta(\hat{a}_1)^r$, that is strictly smaller than $-\beta(b)^r$, since we have overreach when $b \in \left[\frac{\hat{a}_1}{2}, \hat{a}_1\right)$. Similarly, E obtains a payoff of $-\beta(|b - \hat{a}_1|)^r$, strictly smaller than 0. Proposition 3 formalizes this result.

Proposition 3 *Equilibria with executive overreach are Pareto dominated by the (off-path) outcome of $a = b, y = U$.*

The reason for the inefficiency lies in the combination between divergent interests (i.e. a positive bias) and lack of commitment by the court. In this case, a judicial advisory opinion[11] through which the Court could commit to uphold policy b can be Pareto-improving, and the Court may be willing to do so if it anticipates that the alternative outcome is the overreach policy.

Another interesting implication of executive overreach is that, when it arises (i.e. when $p > r$), the welfare effect of judicial review for the average voter is also non-monotonic in the bias. To see this, consider a voter whose payoff function coincides with the payoff function of the court. Therefore,

$$\pi_V(a, y) = -\beta a^r \mathbf{1}_{\{y=U\}} \; - \; \gamma a^p \mathbf{1}_{\{y=S\}}$$

We now compare V's payoffs in the SPNE with $p > r$ with her payoffs in an equilibrium where there is no judicial review, hence where $a = b$. For small b, the voter is strictly better off with judicial review, because $a^* = 0$ and as a consequence she is obtaining the first best payoff. For large b, judicial review is not binding, therefore V is indifferent between having the review or not. For intermediate biases, when executive overreach happens in equilibrium, the voter would be better off without judicial review. Figure 3 illustrates this point, comparing π_V in equilibrium with judicial review (blue line) and without it (red dashed line).

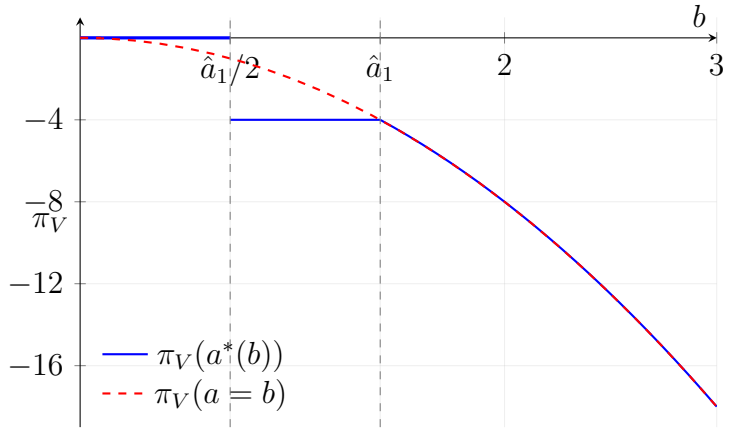


Figure 3: π_V under the equilibrium strategy $a^*(b)$ versus the counterfactual $a = b$. Parameters: $r = 2$, $p = 4$, $\beta = 2$, $\gamma = 1$.

5 Institutional implications

The model highlights a commitment problem on the remedial margin: when reversal becomes sufficiently disruptive after implementation begins, courts rationally shade to-

[11] See Rogers and Vanberg (2002) for a model of judicial advisory opinion with incomplete information.

ward dispositions that preserve the status quo, and executives anticipate this by selecting policies that generate entrenchment. In the overreach regime ($p > r$), judicial review can therefore distort policy *toward* more aggressive implementation rather than disciplining it. This perspective yields several institutional implications.

1. Front-loading and entrenchment are strategic responses to remedy sensitivity. When striking is costly ex post, the executive has an incentive to implement quickly and broadly so as to increase reversal costs. In the language of the model, higher γ lowers the court's upholding cutoff $\hat{a}_1$, shifting the overreach region toward lower executive bias and making implementation possible at lower b.[12] This suggests that observed executive "extremism" under judicial oversight may be concentrated in domains where reliance forms rapidly and unwinding is administratively complex—trade policy provides a clear example through supply-chain reallocation, compliance investments, and refund/administrability burdens once tariffs are collected.

2. The most effective constraints are those that limit pre-review effects. Because the distortion operates through the growth of $D(a)$ before final adjudication, institutions that prevent reliance from accumulating can mitigate overreach without requiring courts to ignore disruption at the remedy stage. Practically, this points to procedural devices such as (i) quicker access to preliminary relief and expedited merits, (ii) automatic or presumptive stays of novel legally contestable measures pending review, and (iii) phased or pilot implementation that caps exposure while legality is litigated. In terms of the model, these tools reduce the mapping from aggressiveness to reversal costs (lowering γ or p for the relevant pre-review window), thereby raising the court's willingness to strike moderately aggressive policies and weakening the executive's incentive to jump to the upholding threshold.

3. "Reversibility design" can substitute for blunt remedial commitment. A different approach is to preserve policy flexibility by reducing the social loss from unwinding. For tariffs, examples include escrow or bonding mechanisms, streamlined refund protocols, and prospective implementation that avoids large retroactive adjustments. More generally, statutory and administrative design choices that make policies easier to unwind reduce disruption conditional on invalidation. In the model, these policies reduce $D(a)$ and shrink the set of parameters under which the court's decision rule becomes entrenchment-based.

4. Non-monotonic welfare effects imply that "more review" is not always better. The voter-welfare comparison in Section 4 implies that the benefits of review

[12]Note, however, that it does not monotonically increase the amount of overreach for every fixed b

are largest when it deters contested policies (small b), and negligible when it does not bind (large b), but can be negative in intermediate cases where it induces overreach. This non-monotonicity cautions against evaluating judicial review solely by average effects and instead suggests focusing empirically on environments where reversal costs are high and convex.

5. When to expect executive overreach. According to the model, executive overreach requires $p > r$, i.e., a relatively steep reversal cost function. Obviously, this does not apply to all policy domains. For example, in the case of surveillance measures the cost of upholding an extreme law may increase especially quickly, hence we should not expect court-motivated overreach. Instead, it is more likely to appear in domains where there are economically relevant adjustment costs and where agents adjust in a forward-looking way. Technological changes or environmental regulations may be relevant domains, on top of trade policy, as compliance to new regulations may require significant, forward-looking investments, and striking down those regulations may lead to significant disruptions.

6 Conclusion

This paper develops a simple formal model of judicial review in which courts face not only legality/precedent concerns but also reversal costs associated with unwinding implemented policies. When reversal costs rise sufficiently with policy aggressiveness, the court's optimal disposition can exhibit an "entrenchment" logic: striking is attractive for mild measures but becomes unattractive for more aggressive ones because disruption dominates. Anticipating this, executives may choose policies that are *more aggressive than their own ideal point* in order to cross the court's upholding threshold. Executive overreach arises for an intermediate range of preference divergence and generates inefficiency: both players would prefer a moderate implement-and-uphold outcome, but the executive distorts implementation to manipulate the court's ex post remedial incentives.

The broader takeaway is that the effectiveness of judicial review depends not only on how courts assess legality, but on whether institutions limit the executive's ability to create costly-to-reverse facts on the ground before legality is finally determined. This shifts attention from the intensity of review to the timing and design of remedies and implementation. Procedural tools that reduce pre-review sunk effects—expedited review, stays, phased implementation, or reversible policy design (e.g. escrow and refund mechanisms for tariffs)—can mitigate strategic entrenchment and restore the disciplining role of judicial oversight. More generally, the analysis suggests that remedy-sensitive judicial behavior can unintentionally encourage aggressive executive action precisely in the environments where unwinding is hardest, a pattern that is likely to be empirically salient in domains such as trade policy and other high-reliance regulatory interventions.

References

Bueno de Mesquita, E. and M. C. Stephenson (2007). Regulatory quality under imperfect oversight. *American Political Science Review 101*(3), 605–620.

Callander, S. and T. S. Clark (2017). Precedent and doctrine in a complicated world. *American Political Science Review 111*(1), 184–203.

Clark, T. S. (2016). Scope and precedent: judicial rule-making under uncertainty. *Journal of Theoretical Politics 28*(3), 353–384.

Delgado-Vega, Á., W. Dziuda, and A. Loeper (2025). The politics of repeal. Working paper. Available at SSRN (abstract ID 4632868).

Fox, J. and M. C. Stephenson (2011). Judicial review as a response to political posturing. *American Political Science Review 105*(2), 397–414.

Fox, J. and M. C. Stephenson (2015). The welfare effects of minority-protective judicial review. *Journal of Theoretical Politics 27*(4), 499–521.

Gely, R. and P. T. Spiller (1990). A rational choice theory of supreme court statutory decisions with applications to the State Farm and Grove City cases. *Journal of Law, Economics, and Organization 6*(2), 263–300.

Gratton, G. and M. Morelli (2022). Optimal checks and balances under policy uncertainty. *International Economic Review 63*(2), 549–569.

Hamilton, A. (1961). Federalist no. 78: The judiciary department. In C. Rossiter (Ed.), *The Federalist Papers*. New York: Signet Classics. Originally published 1788.

La Porta, R., F. Lopez-de Silanes, C. Pop-Eleches, and A. Shleifer (2004). Judicial checks and balances. *Journal of Political Economy 112*(2), 445–470.

Patty, J. W. and I. R. Turner (2021). Ex post review and expert policy making: When does oversight reduce accountability? *The Journal of Politics 83*(1), 23–39.

Rogers, J. R. and G. Vanberg (2002, April). Judicial advisory opinions and legislative outcomes in comparative perspective. *American Journal of Political Science 46*(2), 379–397.

Seidenfeld, M. (1997). Playing games with the timing of judicial review: An evaluation of proposals to restrict pre-enforcement review of agency rules. *Ohio State Law Journal 58*, 85–129.

Shipan, C. R. (2000). The legislative design of judicial review: A formal analysis. *Journal of Theoretical Politics 12*(3), 269–304.

Spiller, P. T. and R. Gely (1992). Congressional control or judicial independence: The determinants of U.S. supreme court labor-relations decisions, 1949–1988. *RAND Journal of Economics 23*(4), 463–492.

Spiller, P. T. and E. H. Tiller (1997). Decision costs and the strategic design of administrative process and judicial review. *The Journal of Legal Studies 26*(2), 347–370.

Stephenson, M. C. (2006). A costly signaling theory of "hard look" judicial review. *Administrative Law Review 58*(4), 753–814.

Tiller, E. H. and P. T. Spiller (1999). Strategic instruments: Legal structure and political games in administrative law. *Journal of Law, Economics, & Organization 15*(2), 349–377.

Turner, I. R. (2017). Working smart and hard? agency effort, judicial review, and policy precision. *Journal of Theoretical Politics 29*(1), 69–96.

Turner, I. R. (2019). Political agency, oversight, and bias: The instrumental value of politicized policymaking. *Journal of Law, Economics, and Organization 35*(3), 544–578.

Vanberg, G. (2001). Legislative–judicial relations: A game-theoretic approach to constitutional review. *American Journal of Political Science 45*(2), 346–361.

Vanberg, G. (2005). *The Politics of Constitutional Review in Germany*. Political Economy of Institutions and Decisions. Cambridge: Cambridge University Press.

A Proofs

Proof of Lemma 1

Proof. Fix $a \geq 0$. The Court chooses U iff

$$\pi_C(a, U) \geq \pi_C(a, S) \quad \iff \quad -\beta a^r \geq -\gamma a^p \quad \iff \quad \beta a^r \leq \gamma a^p.$$

Suppose $a > 0$. Dividing both sides of $\beta a^r \leq \gamma a^p$ by $a^r > 0$ yields

$$\beta \leq \gamma a^{p-r}.$$

Case 1: $p > r$**.** Then $p - r > 0$, so the inequality is equivalent to

$$a^{p-r} \geq \frac{\beta}{\gamma} \quad \iff \quad a \geq \left(\frac{\beta}{\gamma}\right)^{\frac{1}{p-r}} =: \hat{a}_1.$$

Hence, for $a > 0$, the Court chooses U iff $a \geq \hat{a}_1$.

Case 2: $p < r$**.** Then $p - r < 0$, and $\beta \leq \gamma a^{p-r}$ is equivalent to

$$\beta \leq \frac{\gamma}{a^{r-p}} \quad \iff \quad a^{r-p} \leq \frac{\gamma}{\beta} \quad \iff \quad a \leq \left(\frac{\gamma}{\beta}\right)^{\frac{1}{r-p}} =: \hat{a}_2.$$

Hence, for $a > 0$, the Court chooses U iff $a \leq \hat{a}_2$.

Finally, note that if the indifferent point (i.e. $a = \hat{a}_1$ or $a = \hat{a}_2$) leads to $y = U$, an equilibrium exists for every parametric configuration. In the opposite case, a pure strategy SPNE would not exist for some values of the bias, in particular when b is such that E wants to choose a minimal (or maximal) policy leading to U. ■

Proof of Proposition 1

Proof. Assume $p > r$. By Lemma 1, the Court's strategy is: choose U iff $a \geq \hat{a}_1$.

Anticipating this, the Executive's choice problem is:

$$\max_{a \geq 0} \begin{cases} -\beta |a - b|^r, & \text{if } a \geq \hat{a}_1, \\ -\beta b^r, & \text{if } a < \hat{a}_1. \end{cases}$$

Given the tie-breaking rule, if the Executive anticipates $y = S$ she chooses $a = 0$.

Step 1: If $b \geq \hat{a}_1$ then $a^* = b$. If $b \geq \hat{a}_1$, the Executive can choose $a = b$ and induce $y = U$, obtaining payoff $\pi_E(b) = -\beta|b - b|^r = 0$, which is maximal since payoffs are non-positive. Hence $a^* = b$.

Step 2: If $b < \hat{a}_1$ then the Executive compares $a = 0$ and $a = \hat{a}_1$. If $b < \hat{a}_1$, any $a \in (0, \hat{a}_1)$ triggers $y = S$ and yields payoff $-\beta b^r$; by tie-breaking the Executive then prefers $a = 0$ among those. To induce $y = U$, the Executive must choose $a \geq \hat{a}_1$; among such a, she minimizes $|a - b|$ by choosing the smallest feasible a, i.e. $a = \hat{a}_1$, because $b < \hat{a}_1$ implies $|a - b|$ is increasing in a on $[\hat{a}_1, \infty)$. Thus the relevant comparison is between:

$$\pi_E(0) = -\beta b^r \quad \text{and} \quad \pi_E(\hat{a}_1) = -\beta(\hat{a}_1 - b)^r.$$

The Executive chooses $\hat{a}_1$ iff

$$(\hat{a}_1 - b)^r \leq b^r \quad \Longleftrightarrow \quad \hat{a}_1 - b \leq b \quad \Longleftrightarrow \quad b \geq \frac{\hat{a}_1}{2},$$

where we used $b > 0$ and $r > 1$ so that $x \mapsto x^r$ is strictly increasing on $\mathbb{R}_+$.

Putting steps 1 and 2 together:

$$a^* = \begin{cases} 0, & b < \hat{a}_1/2, \\ \hat{a}_1, & b \in [\hat{a}_1/2, \hat{a}_1], \\ b, & b > \hat{a}_1. \end{cases}$$

Given the tie-breaking assumption (Executive chooses $a = 0$ whenever she anticipates $y = S$), this is the unique pure-strategy SPNE. ■

Proof of Proposition 2

Proof. Assume $p < r$. By Lemma 1, for $a > 0$ the Court upholds iff $a \leq \hat{a}_2$.

Thus, the Executive can ensure $y = U$ by choosing any $a \in (0, \hat{a}_2]$, yielding payoff $-\beta|a - b|^r$. Since $r > 1$, the function $a \mapsto |a - b|^r$ is minimized at $a = b$, subject to the feasibility constraint $a \leq \hat{a}_2$.

If $b < \hat{a}_2$: the unconstrained optimum $a = b$ is feasible, so $a^* = b$ and payoff is 0.

If $b \geq \hat{a}_2$: the constrained optimum is the closest feasible point to b, namely $a^* = \hat{a}_2$. This yields payoff $-\beta(b - \hat{a}_2)^r$, which strictly dominates choosing $a = 0$ (payoff $-\beta b^r$) because $b - \hat{a}_2 < b$ and $x \mapsto x^r$ is increasing on $\mathbb{R}_+$.

Hence,

$$a^* = \begin{cases} b, & b < \hat{a}_2, \\ \hat{a}_2, & b \geq \hat{a}_2, \end{cases}$$

and together with the Court's best response this constitutes the unique pure-strategy SPNE under the stated tie-breaking. ■

Proof of Proposition 3

Proof. We interpret "executive overreach" as the region in Proposition 1 (with $p > r$) where the equilibrium choice is $a^* = \hat{a}_1$ even though $b < \hat{a}_1$, i.e. $b \in [\hat{a}_1/2, \hat{a}_1)$. In that region, the Court upholds (since $a^* = \hat{a}_1 \geq \hat{a}_1$), so the equilibrium outcome is $(a, y) = (\hat{a}_1, U)$.

Consider instead the feasible alternative outcome $(a, y) = (b, U)$. Since $b < \hat{a}_1$, under the model's Court decision rule the Court would in fact strike at $a = b$; however, for the welfare/Pareto comparison we can still compare payoffs across outcomes (i.e. across terminal allocations), which is what the paper does in the surrounding text.

Compute payoffs:

Court:

$$\pi_C(\hat{a}_1, U) = -\beta \hat{a}_1^r, \qquad \pi_C(b, U) = -\beta b^r.$$

Because $b \in [\hat{a}_1/2, \hat{a}_1)$ implies $0 < b < \hat{a}_1$ and $x \mapsto x^r$ is strictly increasing on $\mathbb{R}_+$, we have $b^r < \hat{a}_1^r$, hence

$$\pi_C(b, U) > \pi_C(\hat{a}_1, U).$$

Executive:

$$\pi_E(\hat{a}_1, U) = -\beta(\hat{a}_1 - b)^r < 0, \qquad \pi_E(b, U) = -\beta|b - b|^r = 0.$$

Thus $\pi_E(b, U) > \pi_E(\hat{a}_1, U)$.

Therefore, both players are strictly better off under $(a, y) = (b, U)$ than under the

overreach equilibrium outcome $(\hat{a}_1, U)$, so the overreach equilibrium outcome is Pareto dominated by (b, U). ■

B Generalization of the main result

In this appendix, we show that executive overreach exists, as an equilibrium outcome, for more general loss functions than those assumed in the benchmark model.

Define $L(a) := -\pi_C(a)$ as the loss from implementing policy a, and $D(a)$ as the reversal cost. Assume that both functions are continuous and increasing in a. Moreover, assume that $\pi_E(a, b)$ is a continuous, symmetric around $b > 0$ and single peaked payoff function. Define executive overreach as E proposing a policy $a > b$ that is upheld by C.

Proposition 4 *Let $L(a), D(a) : \mathbb{R}_+ \to \mathbb{R}_+$ be continuous and increasing. Assume there exist numbers $0 \leq a^- < \hat{a} < a^+$ such that*

1. *$L(\hat{a}) = D(\hat{a})$;*
2. *$L(a) > D(a)$ for all $a \in (a^-, \hat{a})$;*
3. *$L(a) < D(a)$ for all $a \in (\hat{a}, a^+)$.*

Assume the Executive's payoff $\pi_E(a, b)$ is continuous in a, single-peaked with peak at $b > 0$, and symmetric around b.

Suppose additionally that either *(A) $a^- = 0$,* or *(B) $L(a^-) \leq D(a^-)$. Then there exists a nonempty interval of b for which, in a pure-strategy SPNE, the Executive chooses $a^* = \hat{a} > b$ and the Court upholds (executive overreach). In particular:*

- *In case (A), overreach occurs for all $b \in \left[\frac{\hat{a}}{2}, \hat{a}\right)$.*
- *In case (B), overreach occurs for all $b \in \left[\frac{a^- + \hat{a}}{2}, \hat{a}\right)$.*

Proof. Fix $a > 0$. The Court compares upholding, which yields payoff $-L(a)$, to striking, which yields payoff $-D(a)$. Hence it chooses U iff $L(a) \leq D(a)$.

By assumptions (1)–(3), the Court upholds for all $a \in (\hat{a}, a^+)$ and strikes for all $a \in (a^-, \hat{a})$. At $a = \hat{a}$ it is indifferent; select U.

Case (A): $a^- = 0$. Then the Court strikes for all $a \in (0, \hat{a})$ and upholds for all $a \in [\hat{a}, a^+)$. If the Executive chooses any $a \in (0, \hat{a})$ she anticipates $y = S$; by tie-breaking she chooses $a = 0$ among such actions. Therefore the Executive's relevant comparison is between $a = 0$ (which yields the post-strike outcome) and $a = \hat{a}$ (the smallest upheld action). For

any $b \in (0, \hat{a})$, by symmetry around b and single-peakedness, $\pi_E(a, b)$ is strictly decreasing in $|a - b|$. Thus the Executive prefers $\hat{a}$ to 0 iff

$$|\hat{a} - b| \leq |0 - b| \quad \Longleftrightarrow \quad \hat{a} - b \leq b \quad \Longleftrightarrow \quad b \geq \frac{\hat{a}}{2}.$$

For all $b \in \left[\frac{\hat{a}}{2}, \hat{a}\right)$, the Executive therefore chooses $a^* = \hat{a} > b$, and the Court upholds: executive overreach occurs.

Case (B): $L(a^-) \leq D(a^-)$ (so the Court upholds at a^-). Then the Court upholds at a^-, strikes on $(a^-, \hat{a})$, and upholds at and above $\hat{a}$ (locally). Fix any $b \in (a^-, \hat{a})$. Any $a \in (a^-, \hat{a})$ is struck, so among actions that (locally) induce U the relevant candidates are a^- (a non-overreaching upheld action) and $\hat{a}$ (the smallest upheld action above the strike region). By symmetry and single-peakedness, the Executive prefers the upheld action closer to b. Hence she prefers $\hat{a}$ to a^- iff

$$|\hat{a} - b| \leq |a^- - b| \quad \Longleftrightarrow \quad \hat{a} - b \leq b - a^- \quad \Longleftrightarrow \quad b \geq \frac{a^- + \hat{a}}{2}.$$

Therefore, for all $b \in \left[\frac{a^- + \hat{a}}{2}, \hat{a}\right)$, the Executive chooses $a^* = \hat{a} > b$ and the Court upholds: executive overreach occurs.

In both cases, the specified interval of b is nonempty because its lower endpoint is strictly below $\hat{a}$. ■

Intuitively, Proposition 4 shows that, in order to get overreach in equilibrium, it is enough to have $D(a)$ crossing $L(a)$ from below at least once. In this case, we have that the Court chooses $y = S$ in the interval $(a^-, \hat{a})$ and $y = U$ in the interval $[\hat{a}, a^+)$. Therefore, since $\pi_E(a, b)$ is symmetric around b, overreach will be observed at least when b is between $\frac{\hat{a}+a^-}{2}$ and $\hat{a}$, because E will choose $a = \hat{a} > b$. Note that, once again, the outcome $a = \hat{a}, y = U$ is Pareto-dominated by the (off-equilibrium) outcome $a = b, y = U$.