Why Do Judges Read Statutes?*

Alexander Volokh†

November 6, 2009

Abstract

The standard view that "statutory interpretation matters"—that different methods can "lead to" different results—is hard to square with the standard rational-choice account of judicial decisionmaking. Indeed, under the standard model, it is not obvious why a judge should bother to even read the statute.

I show, within the rational-choice account, how the judge can benefit from reading the statute when the preferences of legislators are uncertain. Doing so shows the judge what policy the legislators agreed to in the past, which gives him clues as to legislators' preferences today. Moreover, different assumptions about how the legislature can react to judicial decisions will alter the judge's decisionmaking.

When an override simply takes the form of a "very large penalty" for the judge, he will in general deviate from his ideal point to the median of the distribution of legislators' preferences, in an effort to avoid being penalized.

However, if an override takes the form of an actual change of policy, not all overrides are *not* created equal. I show that some forms of legislative override do not encourage the judge to read the law, while, surprisingly, other forms of override *do* encourage the judge to *read* the law but can actually *decrease* the extent to which the judge *follows* the law. The structure of a legislative override thus has more complex effects on judicial behavior than has previously been understood.

I also show that, when more than one possible meaning of the statute is available, depending on the method of statutory interpretation is available, some judges may gravitate toward one or the other method depending in part on their own ideology.

Empirical work must therefore take care to distinguish between the political biases of judges who choose a particular interpretive method and the "true nature" of the method, or what opinions using the method would look like if all judges were constrained to use it.

*I am grateful to Ronen Avraham, Bernard S. Black, Robert G. Bone, Seth J. Chandler, David Chichka, John M. Golden, Stefanie A. Lindquist, Angela K. Littwin, Katherine V. Litvak, Daniel B. Rodriguez, Matthew Stephenson, Eugene Volokh, Hanah Metchis Volokh, Vladimir Volokh, and workshop attendees at UT-Austin for their comments and guidance.

†Assistant Professor, Emory Law School, *avolokh@emory.edu*.

Contents

1 Introduction

An intuitive model of statutory interpretation In Volokh (2008), I discussed how methods of statutory interpretation had a "constraining" effect on judges. In our judicial culture, I wrote, judges "have" to write reasoned opinions; the norm of reasoned decisionmaking implies that they incur an "implausibility cost" when they write a less plausible opinion. One way of interpreting the "implausibility cost" is that a statute has a "most plausible point," and a judge's implausibility cost increases as his interpretation deviates from that point.

A judge who only cares about plausibility will thus "comply with the law," i.e., write maximally plausible opinions. A judge who only cares about writing an opinion that embodies his own beliefs will ignore the law and write an opinion that comes as close as possible to his ideal point, subject to the possibility of being overridden by the legislature. The more the judge cares about plausibility, the more "pull" the statute exerts on his decisionmaking. If there are several interpretive methods available, each leading to different "most plausible points," then the judge may choose between methods, picking the one that (according to this process) yields the smallest deviation from his ideal point.

Formalizing the model This paper attempts to formalize that intuitive model.

Here, I do not assume any independent "implausibility cost"; rather, I posit a judge who has only ideological preferences over the set of possible legal interpretations. To the extent he deviates from his own ideal point, it is only to avoid being overridden by the legislature.

Statutes play a special role in this model. If the judge has complete knowledge of where legislative preferences lie, he knows the range within which he will avoid an override, so reading the statute is useless to him. However, if the judge does not know the legislature's preferences, and if legislative preferences have not changed too much since the enactment of the statute, then reading the statute is useful, since it gives the judge some clues as to what policies will and won't be overridden.

The statute at least tells the judge (1) that the leftmost legislator's position lies to the left of the statutory point and (2) that the rightmost legislator's position lies to the right of the statutory point. (I assume, for simplicity, that there are two legislators, a Congressman and a President.) The interval between these two points defines a "no-override" range, since the legislators will never unanimously agree to override a legal interpretation within that range. In particular, the judge also knows that (3) he has a safe harbor if he rules at the point established by the statute.

Thus, the judge doesn't care about complying with the law for its own sake; complying with the law serves an instrumental purpose in helping the judge avoid an override and thus maximize his ideological preferences.

Different behavior under different models of legislative override The judge's behavior depends on what process is followed in case of override. After presenting the simple intuitive model of implausibility cost in Part 2 and discussing its inadequacies as a rational-choice account of judicial decisionmaking, in Part 3 I discuss three different mechanisms:

1. *The "very bad penalty" interpretation.* This is not an override mechanism at all, but a penalty that the legislature can impose on the judge, like a fine or impeachment.

2. *The "closest point" interpretation.* In case of override, the legislature reverts to the closest point in the no-override range.

3. *The "Nash bargaining" interpretation.* In case of override, the legislature reverts to the average between the best feasible point for the Congressman and the best feasible point for the President, where a legislator's "best feasible point" is defined as the point that maximizes his own utility subject to the other legislator's agreement that that point is no worse than the status quo.

Using this simple model, we can prove several results, some intuitive and some not so intuitive. First, if the judge does not read the statute, then:

- Under the "very bad penalty" interpretation, then the judge always deviates somewhat from his ideal point, in the direction of the median of the distribution of legislators' preferences. This is easy to understand, as an effort to avoid being overridden.

- However, once we move to more realistic conceptions of overrides—where an override consists of the legislature replacing the judge's ruling with a new policy that the pivotal legislators all prefer—this intuitive result no longer holds necessarily. In fact, under the simplest reversion rule—the "closest point" rule—*it is not optimal for the judge to deviate at all from his ideal point*.

- However, more "marginally punitive" reversion rules can make the judge deviate from his ideal point. With the "Nash bargaining" rule, I give an example in which the judge deviates from his ideal point toward the median.

Now suppose the judge does read the statute. Does this have any effect?

1. Under the "very bad penalty" interpretation, the judge moves even closer to the statutory point.

2. However, under the "closest point" rule, the judge continues to always rule at his ideal point. *Reading the statute does not make the judge any more likely to deviate from his ideal point.*

3. Most surprisingly, under more punitive reversion rules, the effects of reading the statute can be unpredictable. I give two examples where, under the "Nash bargaining" rule, reading the statute can make the judge rule closer to the statute, *or could make him rule further away from it*.

It is clear, then, that the existence of statutes has a complicated—and sometimes surprising—relationship to judicial decisionmaking.

The effect of different methods of statutory interpretation Now suppose that there is more than one method of statutory interpretation available—say textualism and intentionalism.

- *Textualism* is an interpretive theory according to which the meaning of a statute is determined by the *semantic* meaning of the terms. In other words, the objective meaning of the text controls. This means the meaning that would be understood by a reasonable, educated person familiar with the law, legal language and legal usage, including how similar terms are used elsewhere in the law and how these and similar terms have been interpreted in relevant legal precedents. According to textualism, evidence of the subjective intent of the legislators cannot be used to help interpret an ambiguous statute.

- *Intentionalism* is an interpretive theory according to which the meaning of a statute is determined by the *intent* of the legislators that enacted it. In other words, the subjective meaning of the text controls. This meaning *may* be revealed by the text, and typically if the text is unambiguous, the inquiry ends there; but in case of ambiguity, intentionalism holds that the meaning can be determined with the help of sources such as legislative committee reports and the text of legislative debates.

- There may be other theories, for instance *purposivism*, according to which the meaning of the statute depends on the problem that the statute was enacted to address. For an overview of different methods of statutory interpretation, see Eskridge, Frickey, and Garrett (2007).

In most cases, it does not matter which method of statutory interpretation one follows: They all lead to the same result. However, occasionally, different methods of statutory interpretation lead to different "most plausible points." This can create discontinuities in the apparent distribution of judges. Some judges will "gravitate" toward one statutory point, while others will move toward another. Depending on where the judge's ideal point is, he may move toward one point or another.

This has implications for the empirical study of interpretive methods. If there were only one interpretive method, all judges would move toward that method's most plausible point, so a smooth distribution of judges will yield a smooth distribution of judicial decisions. But with several methods, different judges move toward different points, creating the illusion of stark ideological gaps between interpretive methods even if those gaps are small or nonexistent. As I wrote in Volokh (2008), commenting specifically on the conventional wisdom about textualism:

> The statutory interpretation literature is teeming with claims about textualism—its supposed political bias, its tendency to produce congressional overrides, and its tendency to find plain meaning. . . But this conventional wisdom may be mistaken: It fails to take into account that the textualism we observe in written judicial opinions may be an unrepresentative sample of textualist analysis as a whole. . . .
>
> . . . [I]ndividual judges—who today have broad choice among interpretive methods—will tend to select the interpretive method that, other things being equal, minimizes the extent to which they must deviate from their preferred outcomes. This self-selection effect can seriously mislead observers as to the nature of different interpretive methods.
>
> To illustrate, suppose that textualism and intentionalism lead to almost identical ranges of possible policy results, with the textualism range being only slightly more conservative. If lawmakers adopted a Federal Rule of Statutory Interpretation mandating one method or another, or if the Supreme Court took the advice of the House

> of Lords and mandated a method by judicial fiat, the resulting distributions of judicial opinions would be nearly equivalent. But in a regime of free choice, conservatives would tend to choose textualism and liberals would tend to choose intentionalism. This would substantially exaggerate the political differences between the methods: We would observe only the most conservative possible textualist opinions and the most liberal possible intentionalist opinions.
>
> . . . In short, many statements about *textualism* may really only be statements about *textualists*.

Directions for further research These are the first steps in a larger project. The next steps are:

- To refine the behavioral assumptions of judges and legislators:
  First, as I have remarked above, in the current model, observing a statute with a point M tells the judge that the leftmost legislator is to the left to M and the rightmost legislator is to the right of M. But if we know what bargaining process legislators use to enact statutes, we may be able to extract more information than that.
  Second, when there is more than one interpretive method available, I assume that this is inadvertent because a legislature would always prefer to enact a fixed point. But ambiguity as to statutory meaning may be the result of a strategic choice.
  Third, I have assumed that legislators do not enact their statute in anticipation of the judge's behavior. But suppose the legislators knew that judges were extremists on one side of the political spectrum; then they might enact a statute that bent over in the other direction—even if they did not prefer that point—knowing that the judge would moderate the statute toward his own preferences.

- To discuss how a statutory method can be more or less "constraining." Intuitively, it seems as though some interpretive methods are highly determinate while others are less so; debates over textualism, intentionalism, and purposivism often engage this question. The question in this model is what rational-choice assumptions can make methods more or less constraining.

- To discuss the implications for empirical study of judicial decisionmaking. Because judges can choose which method they follow from case to case, an empirical method appropriate for self-selection may be appropriate.

2 The intuitive model of implausibility cost and its problems

In this Part, I present the intuitive model based on my discussion in Volokh (2008), where "ideological utility" is traded off against "implausibility cost."

In Part 3, I drop the assumption of an independent implausibility cost, and see whether compliance with the statute can be made to emerge from purely ideological preferences.

2.1 Setup

Consider a judge deciding a case in which he has to interpret a statute. The decision in the case can be represented as a point x along a unidimensional policy spectrum. The judge has one interpretive method to choose from.

Let the judge's utility be:

$$v(x) \equiv u(x) - c(x). \tag{1}$$

Ideological utility The first component of judicial utility, $u(x)$, is "ideological utility" (or "agenda-based utility"), or the pleasure the judge gets from implementing his political agenda.

Assume the judge has an ideal point is J, and $u(x)$ is continuous and differentiable, has a maximum at J, and is concave (i.e., $u'(x) > 0$ for all $x < J$, $u'(x) < 0$ for all $x > J$, $u''(x) < 0$ for all x).

Implausibility cost The second component of judicial utility, $c(x)$, is "implausibility cost." Any interpretive method, when applied to a given statutory provision, is assumed to yield some "most plausible point." Since there is one interpretive method here, there is a single most plausible point, which we denote $x = M$. Without loss of generality, assume $J < M$.

A judge can render a decision different from $x = M$ in a case, but in doing so, he writes a less plausible point and thus incurs $c(x)$. Assume $c(x)$ is continuous and differentiable, has a minimum at M (i.e., $c'(x) < 0$ for all $x < M$, and $c'(x) > 0$ for all $x > M$), and is convex (i.e., $c''(x) > 0$ for all x).

2.2 The mutual pull of ideology and plausibility

Proposition 1 *Denote the judge's choice as $x^*(J, M, c)$.*

(a) Under these assumptions, $x^*(J, M, c) \in (J, M)$.

(b) A "more determinate" interpretive method "constrains" the judge more. That is, given $\alpha > 1$, we have:

$$\begin{cases} |x^*(J, M, \alpha c) - M| < |x^*(J, M, c) - M|, \\ \lim_{\alpha \to \infty} x^*(J, M, \alpha c) = M. \end{cases} \tag{2}$$

(c) Similarly, a "less determinate" interpretive method "constrains" the judge less. That is, given $\alpha \in (0, 1)$, we have:

$$\begin{cases} |x^*(J, M, \alpha c) - J| < |x^*(J, M, c) - J|, \\ \lim_{\alpha \to 0} x^*(J, M, c) = J. \end{cases} \tag{3}$$

Proof. (a) The judge chooses x to maximize $v(x)$. The derivative of $v(x)$ is:

$$v'(x) \equiv u'(x) - c'(x). \tag{4}$$

Evaluating this derivative at J, we obtain:

$$v'(J) = u'(J) - c'(J) > 0, \tag{5}$$

and since:

$$v''(x) \equiv u''(x) - c''(x) < 0 \text{ for all } x, \tag{6}$$

we have:

$$x^*(J, M, c) > J. \tag{7}$$

Similarly, evaluating the derivative at M, we obtain:

$$v'(M) = u'(M) - c'(M) < 0, \tag{8}$$

and since $v''(x) < 0$ (by (6)), we have:

$$x^*(J, M, c) < M. \tag{9}$$

(b, c) Denote by $v_\alpha(x)$ the judicial utility function when implausibility cost is $\alpha c(x)$. The judge chooses x to maximize $v_\alpha(x)$. The derivative of $v_\alpha(x)$ is:

$$v'_\alpha(x) \equiv u'(x) - \alpha c'(x). \tag{10}$$

Evaluating this derivative at $x^*(J, M, c)$, we obtain:

$$\begin{aligned} v'_\alpha(x^*(J,M,c)) &\equiv u'(x^*(J,M,c)) - \alpha c'(x^*(J,M,c)) \\ &\equiv u'(x^*(J,M,c)) - c'(x^*(J,M,c)) \\ &\quad -(\alpha-1)c'(x^*(J,M,c)) \\ &= -(\alpha-1)c'(x^*(J,M,c)), \end{aligned} \tag{11}$$

which is positive for $\alpha > 1$ and negative for $\alpha < 1$. Since it is clear that $v''_\alpha(x) < 0$ for all x, we thus have:

$$\begin{cases} x^*(J,M,\alpha c) > x^*(J,M,c) \text{ for } \alpha > 1, \\ x^*(J,M,\alpha c) > x^*(J,M,c) \text{ for } \alpha \in (0,1). \end{cases} \tag{12}$$

Now evaluate the derivative at any point $x_0 \in (J, M)$:

$$v'_\alpha(x_0) \equiv u'(x_0) - \alpha c'(x_0), \tag{13}$$

which is positive for large enough α and negative for α close enough to 0. Thus, for large enough α, $v'_\alpha(x) = 0$ for some $x > x_0$, and thus $x^*(J, M, \alpha c) > x_0$. Therefore, $\lim_{\alpha\to\infty} x^*(J, M, \alpha c) = M$. Similarly, for α close enough to 0, $v'_\alpha(x) = 0$ for some $x < x_0$, and thus $x^*(J, M, \alpha c) < x_0$. Therefore, $\lim_{\alpha\to 0} x^*(J, M, \alpha c) = J$. ■

Example 1

Suppose $u(x) \equiv -(x-J)^2$ and $c(x) \equiv \alpha(x-M)^2$. Then the judge chooses x to maximize $v(x) \equiv -(x-J)^2 - \alpha(x-M)^2$. His first-order condition is:

$$v'(x^*) \equiv -2(x^*-J) - 2\alpha(x^*-M) \equiv -2(1+\alpha)x^* + 2J + 2\alpha M = 0, \quad (14)$$

which implies:

$$x^* = \frac{J+\alpha M}{1+\alpha} = \frac{1}{1+\alpha}J + \frac{\alpha}{1+\alpha}M = \frac{1}{1+\alpha}J + \left(1 - \frac{1}{1+\alpha}\right)M, \quad (15)$$

and since $v''(x) \equiv -4 < 0$, x^* is a maximum. We can see immediately that x^* is a weighted average of J and M with weights 1 and α, respectively, and that therefore, x^* is between J and M for any positive weight α, approaches M as α increases, and approaches J as α decreases.

2.3 Separation under a choice regime

Now suppose that there are two interpretation, textualism and intentionalism, each with its most plausible point, M_1 and M_2, respectively, and each with its own implausibility cost, $c_1(x)$ (having a minimum at M_1) and $c_2(x)$ (having a minimum at M_2). Define:

$$v_i(x) \equiv u(x) - c_i(x) \text{ for } i \in \{1,2\}. \quad (16)$$

Without loss of generality, assume that $M_1 < M_2$.

Now the judge maximizes:

$$v(x) \equiv \max_i \{v_i(x)\}. \quad (17)$$

That is, either he "uses" textualism (i.e., writes an opinion using textualist reasoning), in which case the previous section dictates that he will deviate from J toward M_1, or he uses intentionalism (i.e., writes an opinion using intentionalist reasoning), in which case he deviates from J toward M_2. It is straightforward to show that there is a threshold $\hat{J}$ such that the judge deviates toward M_1 when $J < \hat{J}$ and deviates toward M_2 when $J > \hat{J}$. This self-selection effect can mislead the casual observer, who may see an apparent ideological gulf between the two interpretative methods when in fact—if either method were mandated for the entire judiciary—the ideological difference would be very slight or even nonexistent. (See Volokh (2008, pp. 803–19).)

2.4 Problems with this model: The ideological model with full information

The main problem with this model is that it is ad hoc; while the idea that judges have ideological preferences is quite standard (see, e.g., Eskridge & Frickey

(1994), p. 33; Kennedy (1986); but see Posner (1993), pp. 3, 13–15), it is unclear why judges would also care about the plausibility of their opinions.

One possibility would be judges' internalized responsibility to produce reasoned decisions. This is possible, but for a rational-choice model, one would prefer not to simply assume a preference for producing reasoned decisions. It would be better to make this preference emerge from underlying preferences, for instance the ideologically based preference to not be overridden, which stems from the same source as the ideological preference that makes judges want to rule at their ideal point.

Moreover, the view that judges seek plausibility because they fear being overriden by the legislature, or being overruled by higher courts (which themselves fear legislative overrides), may sit uneasily with the following insights from the standard theory of legislative overrides (see, e.g., Marks (1989), pp. 10–32; Ferejohn & Weingast (1992), pp. 574–76).

Assume there are two unitary actors, whom I will generically call "legislators," whose assent is required to pass a law: a Congressman whose ideal point is C and a President whose ideal point is P. Without loss of generality, assume that $C < P$.

(For simplicity, I ignore the division of Congress into a separate Senate and House of Representatives, both of which have to assent to legislation. I also ignore the collective nature of Congress and assume, again for simplicity, that, on this one-dimensional spectrum, one can limit one's attention to the median Congressman. Finally, I ignore the possibility of veto overrides, where Congress can pass a law without the President's consent.)

The judge now only has ideological preferences $u(x)$, which behave as in Section 2.1.

As long as the judge chooses $x \in [C, P]$, he will be immune to legislative override: Any move away from x would move further from either C or P, and therefore at least one necessary part would not assent to the change.

If the judge chooses $x \notin [C, P]$, then there exists $\hat{x} \in [C, P]$ that is closer to both C and P. For example, if $x < C$, then $\hat{x} = C$ is closer to both C and P; and likewise for $\hat{x} = P$ if $x > P$. Therefore, the legislators would unanimously assent to replacing x with $\hat{x}$. This does not mean that they would actually adopt $\hat{x}$ rather than some other point that they each prefer to x—adopting $\hat{x}$ corresponds to the "closest point" reversion rule that I discuss in Part 3—but it does prove the existence of at least one point that is unanimously preferred to x.

Thus, if $x \notin [C, P]$, suppose the legislature overrides x and establishes some new point. Call this point $\hat{x}^C(x; C, P)$ if $x < C$, and $\hat{x}^P(x; C, P)$ if $x > P$. Both of these are in the interval $[C, P]$.

It is clear that the judge cannot do better than to choose:

$$x^*(C, P) \equiv \begin{cases} J \text{ if } J \in [C, P], \\ C \text{ if } J < C, \text{ and} \\ P \text{ if } J > P. \end{cases} \tag{18}$$

In this model, there is never a legislative override. Moreover, a judge never benefits by reading the statute or otherwise consulting the law. Methods of statutory interpretation and notions of plausibility play no role.

However, this seems to conflict with the observation that judges do (at least sometimes) consult the statute, and that overrides (at least sometimes) happen. The models in the rest of this paper therefore try to build up a rational-choice model of how interpretive methods can constrain, and how legislatures can override in equilibrium, even when judges have solely ideological preferences.

3 The ideological model with an uncertain override range

This section complicates the previous model, by showing what the judge will do when he is uncertain as to the location of C and P. Wherever he rules, his ruling will be upheld if it is between C and P, and overridden if it is outside of that range. Under this family of models, the judge often benefits from consulting the statute, and because of his uncertainty as to legislative preferences, overrides sometimes happen.

Judicial utility The judge's behavior depends on his utility function: As in Part 2.4, I assume that he has an ideological utility function $u(x)$, defined on the set of possible policy positions and behaving as in Part 2.1; moreover, I assume that he is an expected utility maximizer.

Legislators' distribution The judge's behavior also depends on the distribution of legislators' preferences. Suppose that all legislators are drawn from a common legislative distribution L with density function $f(x)$, which we assume has full support. Call the two legislators L_1 and L_2, and define $C \equiv \min\{L_1, L_2\}$ and $P \equiv \max\{L_1, L_2\}$. Note the joint distribution and density functions of C and P, defined for $c < p$:

$$\begin{aligned}
F_{C,P}(c,p) &\equiv \Pr(C < c \cap P < p) \equiv \Pr(C < c \cap P \in (C,p)) \\
&\equiv 2\Pr(L_1 < c \cap L_2 \in (L_1,p)) \equiv 2\int_{-\infty}^{c}\int_{x}^{p} f(x)f(y)dydx \\
&\equiv F^2(p) - (F(p) - F(c))^2, \\
f_{C,P}(c,p) &\equiv \frac{\partial^2}{\partial c \partial p} F_{C,P}(c,p) \equiv 2f(c)f(p).
\end{aligned} \tag{19}$$

If the judge rules at a point x, he will avoid an override provided $x \in [C, P]$. If all he knows is the general distribution of legislators' preferences, this happens with probability:

$$\Pr(x \in [C,P]) \equiv \Pr(C < x \cap P > x) \equiv 2F(x)(1 - F(x)). \tag{20}$$

Why read the statute? Why would the judge want to read the statute, if the probability of override only depends on his position relative to that of the pivotal legislators?

Reading the statute may be useful to the judge because it offers him a way of finding a clue as to the location of C and P. Suppose again there is a single interpretive method, with a most plausible point M. It is reasonable to suppose that—given constant legislative preferences—$C < M$ and $P > M$, or else the legislators would not have agreed to enact M. (Note, though, that this assumes that the legislators are not acting strategically with respect to the judge. This may be an undesirable assumption.)

Thus, at least the judge can avoid an override by ruling at $x = M$. Moreover—if $C < M$ and $P > M$ are the only facts one uses—one can derive new, truncated distributions of C and P, which will aid the judge in gauging the costs and benefits of deviating from M. (If we know something more about legislative bargaining, we can do even better, but suppose that we only use the fact that M is within the no-override range.)

Given that $C < M$ and $P > M$, we can derive the new distribution and density function of C for $x < M$:

$$\begin{aligned} \Phi_C(x) &\equiv \Pr(C < x | C < M) \equiv \frac{F(x)}{F(M)}, \\ \phi_C(x) &\equiv \frac{d}{dx}\Phi_C(x) \equiv \frac{f(x)}{F(M)}. \end{aligned} \tag{21}$$

Similarly:

$$\begin{aligned} \Phi_P(x) &\equiv \Pr(P < x | P > M) \equiv \frac{F(x) - F(M)}{1 - F(M)}, \\ \phi_P(x) &\equiv \frac{d}{dx}\Phi_P(x) \equiv \frac{f(x)}{1 - F(M)}, \end{aligned} \tag{22}$$

though we will typically not use these functions, as we will assume, without loss of generality, that $J < M$, so that we only need to worry about C.

From the C side, the judge can avoid an override as long as $x > C$, that is, with a probability $\Phi_C(x)$. Of course, this assumes that the judge does not stray to the right of M, but it is clear that, if $J < M$, he will limit himself to the range $x \in [J, M]$. Ruling at $x = J$ is at least as good as ruling at any $x < J$, since a lower J increases the chance that the judge will incur a penalty while simultaneously leading to a worse policy outcome for the judge. For analogous reasons, ruling at $x = M$ is at least as good as ruling at any $x > M$.

Override assumptions We must also make assumptions about what exactly the legislature will do in case of an override.

One assumption we can make is to ignore the "override" aspect entirely, and assume instead that the legislature merely penalizes the judge, essentially imposing on him a constant low utility $\hat{u} < 0$.

However, overrides are not a constant penalty, but rather the adoption of some alternative policy point. It is therefore important to see how the judge acts under more realistic override rules. Therefore, if we take overrides seriously, we should specify what the legislature does if the judge's ruling is outside of $[C, P]$—that is, what form $\hat{x}^C(x; C, P)$ or $\hat{x}^P(x; C, P)$ (as defined in Part 2.4) take. I consider two possibilities:

- the "closest point" reversion rule, where the legislature moves the judge's point to the closest point within the interval $[C, P]$; and
- the "Nash bargaining" reversion rule, where Congress and the President determine the override point by Nash bargaining.

Summary of results The results are interesting:

- Under the "very large penalty" interpretation, the judge deviates from J toward the median of f. After the judge reads the statute, he deviates still further in the direction of the statutory point. This is very intuitive.
- Under the "closest point" override rule, *the judge does not deviate from his ideal point at all*. Whether before or after he reads the statute, he stays at his ideal point, and does not deviate toward either the median of f or toward the statutory point M. *Neither the statute nor the possibility of overrides makes him comply with the law or try to please legislators.* This seems surprising, but is actually sensible: While this reversion rule is slightly punitive, it is not *marginally* punitive.
- The "Nash bargaining" rule is more punitive and does lead to some deviation from the ideal point. I provide an example that shows how the judge can deviate in the direction of the median of f. After the judge reads the statute, the results are surprising. I give two examples where the judge deviates from his ideal point toward the statutory point. However, the existence of the statute does not always pull the judge toward the statute. In fact, *the judge's optimum may move closer to the statutory point after he has read the statute, or it may move further away.*

The next sections discuss each of these three possible legislative responses in turn.

3.1 The "very large penalty" interpretation

As a simplified illustration, suppose that the legislature imposes a very large penalty on the judge, which gives the judge a constant utility $\hat{u} < 0$ which is "extremely low" (i.e., "extremely negative"). Define $\hat{v}(x)$ to be the judge's ideological utility $u(x)$ (as defined in Part 2.4) if he is not overridden, and to be $\hat{u}$ if he is overridden. (We could, alternatively, let the judge "keep" his $u(x)$ with 100% probability, but additionally impose a penalty $\hat{u}$ if he is overridden.

This gives similar results. The version of $\hat{v}(x)$ I am using may be less realistic, but it has the virtue of mirroring the "true override" rules, where the judge doesn't get to "keep" the consequences of his rule if he's overridden.)

The penalty $\hat{u}$ could be impeachment, censure, or whatever else the judge dislikes. One thing it is *not* is an actual statutory reversion point, which would be more realistic. It doesn't make sense to thing of the $\hat{u}$ as a "very bad reversion point": Such an interpretation might imply that the legislature would choose a point that is disastrous for all involved, including the legislators themselves. Moreover, $\hat{u}$ is a constant, and may be unrealistic (beyond the unrealism already inherent in the $\hat{u}$ interpretation) to think that the legislature would choose a reversion point that doesn't depend on the precise value of x, only on the fact that x is outside of the range $[C, P]$.

The use of the "very large penalty" thus serves as an illustration to generate intuition.

Without reading the statute First, we can derive the judge's behavior when there is no statute to read—or if the judge does not bother to read the statute.

We have already derived in (20) the probability that the judge avoids an override. The judge's expected utility is thus:

$$\begin{aligned} E\hat{v}(x) &\equiv 2F(x)(1-F(x))u(x) + (1-2F(x)(1-F(x)))\hat{u} \\ &\equiv 2F(x)(1-F(x))(u(x)-\hat{u}) + \hat{u}. \end{aligned} \tag{23}$$

The first-order condition is:

$$\begin{aligned} \frac{d}{dx}E\hat{v}(x^*) &\equiv 2F(x^*)(1-F(x^*))u'(x^*) + 2f(x^*)(1-2F(x^*))(u(x^*)-\hat{u}) \\ &= 0. \end{aligned} \tag{24}$$

The second derivative is:

$$\begin{aligned} \frac{d^2}{dx^2}E\hat{v}(x) &\equiv 4f(x)(1-2F(x))u'(x) + 2F(x)(1-F(x))u''(x) \\ &\quad + [2f'(x)(1-2F(x)) - 4f^2(x)](u(x)-\hat{u}). \end{aligned} \tag{25}$$

All terms of this expression are negative (for the negativity of the first term, see the proof of Proposition 2 below) except for the $2f'(x)(1-2F(x))$ term, whose sign depends on that of $f'(x)$. Therefore, not all solutions of the first-order condition are necessarily maxima. However, it is clear that $E\hat{v}(x^*)$, being continuous and bounded above (by $u(J)$), has a maximum, and that the first-order condition is satisfied at this maximum.

Proposition 2 *The judge chooses x^* between J and L_M, where L_M is the median of the distribution of L (i.e., $F(L_M) = \frac{1}{2}$). He chooses $x^* = J$ or $x^* = L_M$ iff $L_M = J$. Moreover, $\lim_{\hat{u}\to-\infty} x^* = L_M$.*

Proof. (a) First, we prove that $x^* = J$ iff $L_M = J$.

If $x^* = J$, then $u'(x^*) = 0$, and (24) implies that $F(x^*) = \frac{1}{2}$, which means $L_M = J$.

If $L_M = J$, then $F(J) = \frac{1}{2}$; (24) is satisfied at $x = J$ and the second derivative is negative; therefore, $x = J$ is a maximum.

(b) Next, we prove that $x^* = L_M$ iff $L_M = J$.

If $x^* = L_M$, then $F(x^*) = \frac{1}{2}$, and (24) implies that $u'(x^*) = 0$, which means $x^* = J$.

If $L_M = J$, then the proof of (a) implies that $x^* = L_M$.

(c) Now we prove that, when $L_M \neq J$, x^* is strictly between J and L_M.

Observe that (24) reduces to:

$$\frac{1 - 2F(x^*)}{u'(x^*)} = -\frac{F(x^*)(1 - F(x^*))}{f(x^*)(u(x^*) - \hat{u})} < 0, \tag{26}$$

so the sign of $1 - 2F(x^*)$ must be the opposite of the sign of $u'(x^*)$.

Therefore, $x^* > J$ (i.e., $u'(x^*) < 0$) iff $F(x^*) < \frac{1}{2}$ (i.e., $x^* < L_M$).

Conversely, $x^* < J$ (i.e., $u'(x^*) > 0$) iff $F(x^*) > \frac{1}{2}$ (i.e., $x^* > L_M$).

Thus, x^* is strictly between J and L_M.

(d) Now we prove that $\lim_{\hat{u} \to -\infty} x^* = L_M$. Considering (26), we see that the right-hand side approaches 0 as $\hat{u} \to -\infty$, since all the other terms in the expression are bounded, and $f(x^*)$ is bounded away from 0, on the interval $[J, L_M]$. Therefore, the left-hand side must also approach 0. Because $u'(x^*)$ is bounded on the interval $[J, L_M]$, this means that $1 - 2F(x^*)$ must approach 0, which means $F(x^*)$ must approach $\frac{1}{2}$, which means that x^* must approach L_M.
■

In other words, the judge deviates from his ideal point J in the direction of the median of the distribution of the legislator's position, the more so (eventually) as the override penalty increases.

Example 2

Suppose that $u(x) \equiv -(x - J)^2$, and that the legislative position is distributed uniformly on $[0, 1]$, so $f(x) \equiv 1$ and $F(x) \equiv x$ for $x \in [0, 1]$, and $L_M = \frac{1}{2}$. The judge's expected utility is:

$$\begin{aligned} E\hat{v}(x) &\equiv -2x(1 - x)(x - J)^2 + \hat{u}(1 - x)^2 + \hat{u}x^2 \\ &\equiv (-2x + 2x^2)(x^2 - 2xJ + J^2) + \hat{u}(1 - 2x + 2x^2) \\ &\equiv 2x^4 - 2x^3(1 + 2J) + 2x^2(2J + J^2 + \hat{u}) - 2x(J^2 + \hat{u}) + \hat{u}, \end{aligned} \tag{27}$$

and so his first-order condition is:

$$\begin{aligned} \frac{d}{dx}E\hat{v}(x^*) &\equiv -2(1 - 2x^*)((x^* - J)^2 + \hat{u}) - 4x^*(1 - x^*)(x^* - J) \\ &\equiv 8x^{*3} - 6x^{*2}(1 + 2J) + 4x^*(2J + J^2 + \hat{u}) - 2(J^2 + \hat{u}) \\ &= 0. \end{aligned} \tag{28}$$

Since $\frac{d}{dx}E\hat{v}(J) = 2\hat{u}(2J-1)$, it is clear that $x^* = J$ iff $J = \frac{1}{2}$. Since $\frac{d}{dx}E\hat{v}(J) = 4\hat{u}\left(J - \frac{1}{2}\right)$ while $\frac{d}{dx}E\hat{v}\left(\frac{1}{2}\right) = J - \frac{1}{2}$, it is clear that the derivative of $E\hat{v}(x)$ has opposite signs at J and $\frac{1}{2}$ and therefore has a zero between these. Moreover, as $\hat{u} \to -\infty$, $\frac{d}{dx}E\hat{v}(x^*)$ is dominated by the terms $4x\hat{u} - 2\hat{u} = 2\hat{u}(2x-1)$; thus, as $\hat{u} \to -\infty$, $x^* \to \frac{1}{2} = L_M$. Note, finally, that the second derivative of $E\hat{v}(x)$ is:

$$\begin{aligned} \frac{d^2}{dx^2}E\hat{v}(x) &\equiv 24x^2 - 12x(1+2J) + 4(2J + J^2 + \hat{u}) \\ &\equiv 4[6x^2 - 3x(1+2J) + 2J + J^2 + \hat{u}], \end{aligned} \tag{29}$$

which is negative for all $J \in [0,1]$ and $\hat{u} < 0$; so the solutions are indeed maxima.

After reading the statute Denote the judge's new utility as $\hat{v}^M(x)$, where the M superscript stands for the statute's "most plausible point" M. After reading this statute, the judge's expected utility becomes:

$$\begin{aligned} E\hat{v}^M(x) &\equiv \frac{F(x)}{F(M)}u(x) + \left(1 - \frac{F(x)}{F(M)}\right)\hat{u} \\ &\equiv \frac{F(x)}{F(M)}(u(x) - \hat{u}) + \hat{u}. \end{aligned} \tag{30}$$

The judge then chooses $\tilde{x}$ to maximize $E\hat{v}^M(x)$:

$$\frac{d}{dx}E\hat{v}^M(\tilde{x}) \equiv \frac{f(\tilde{x})}{F(M)}(u(\tilde{x}) - \hat{u}) + \frac{F(\tilde{x})}{F(M)}u'(\tilde{x}) = 0. \tag{31}$$

Similarly, if we had a judge with $J > M$, his expected utility would be:

$$E\hat{v}^M(x) \equiv \frac{1 - F(x)}{1 - F(M)}(u(x) - \hat{u}) + \hat{u}, \tag{32}$$

though we will not use this for now since we are assuming $J < M$.

Note that the first-order condition in (31), with equality, is correct as long as $\tilde{x} \leq M$; if (31) is satisfied at a value above M, then the optimum is a corner solution at $\tilde{x} = M$. Note also that (unless we have a corner solution at M), the judge's optimum is actually independent of M (since the $F(M)$ denominators are irrelevant). (As before, the second derivative, which is:

$$\frac{d^2}{dx^2}Ev^M(x) \equiv \frac{f'(x)(u(x) - \hat{u}) + 2f(x)u'(x) + F(x)u''(x)}{F(M)}, \tag{33}$$

is not guaranteed to be negative, as the first term depends on the sign of $f'(x)$. However, it is clear that either there is a maximum at which the first-order condition is satisfied, or there is a corner solution at $\tilde{x} = M$.)

Note also that the derivative is positive at $x = J$, so it is not optimal for him to stay at his ideal point. Thus, the solution will be some $\tilde{x} \in (J, M]$.

If, before reading the statute, the judge would have chosen some $x^* \leq J$ or some $x^* > M$, the discussion above indicates that he would do better to switch to some point within $(J, M]$; so in those cases, reading the statute moves him in the direction of M. (I cannot rule out that, if $x^* > M$, the judge might overshoot M and choose some $\tilde{x} < M$.)

Now, let us consider the case where $x^* \in [J, M]$. Does the statute make him move to some point closer to M? To answer this question, consider $\frac{d}{dx}E\hat{v}^M(x^*)$, that is, the derivative of the judge's utility at the point he would have chosen if he were ignorant of the statute.

We know that, when $F(x^*) = \frac{1}{2}$, we have $x^* = J$. Then the derivative of expected utility, evaluated at the previous optimum, x^*, is:

$$\frac{d}{dx}E\hat{v}^M(x^*) = \frac{d}{dx}E\hat{v}^M(J) = \frac{f(J)}{F(M)}(u(J) - \hat{u}) > 0. \tag{34}$$

And when $F(x^*) \neq \frac{1}{2}$, (24) reduces to:

$$u(x^*) - \hat{u} = -\frac{F(x^*)(1 - F(x^*))}{f(x^*)(1 - 2F(x^*))}u'(x^*). \tag{35}$$

The derivative, (31), then becomes:

$$\begin{aligned} \frac{d}{dx}E\hat{v}^M(x^*) &= \frac{u'(x^*)F(x^*)}{F(M)}\left(1 - \frac{1 - F(x^*)}{1 - 2F(x^*)}\right) \\ &= -\frac{u'(x^*)F^2(x^*)}{F(M)(1 - 2F(x^*))}, \end{aligned} \tag{36}$$

which is positive, since we know from Proposition 2 that $u'(x^*)$ and $1 - 2F(x^*)$ have opposite signs.

So $\frac{d}{dx}E\hat{v}^M(x^*) > 0$ in all cases when $x^* \in [J, M]$. Therefore, the judge moves toward M. The judge's optimum before having read the statute is no longer optimal once he has read the statute. In other words, he benefits from reading the statute.

Example 3

Consider again the example where L is uniformly distributed on $[0, 1]$, and $u(x) \equiv -(x - J)^2$. The judge's expected utility is:

$$E\hat{v}^M(x) \equiv -\frac{x(x - J)^2}{M} + \left(1 - \frac{x}{M}\right)\hat{u}, \tag{37}$$

and so his first-order condition is:

$$\frac{d}{dx}E\hat{v}^M(\tilde{x}) \equiv \frac{-3\tilde{x}^2 + 4J\tilde{x} - (J^2 + \hat{u})}{M} = 0, \tag{38}$$

which has solutions at $\tilde{x} = \frac{2}{3}J \pm \frac{1}{3}\sqrt{J^2 - 3\hat{u}}$. However, the second derivative of $E\hat{v}^M(x)$ is:

$$\frac{d^2}{dx^2}E\hat{v}^M(x) \equiv \frac{-6x + 4J}{M}, \tag{39}$$

which is negative for $x > \frac{2}{3}J$. So the only maximum is at $\tilde{x} = \min\left\{\frac{2}{3}J + \frac{1}{3}\sqrt{J^2 - 3\hat{u}}, M\right\}$ (since there will be a corner solution at M if the first-order and second-order conditions are satisfied at some point to the right of M).

Let us check whether, when $x^* \in (J, M)$, the optimum moves to the right after the judge reads the statute—that is, whether $\tilde{x} > x^*$. We do so by establishing that $\frac{d}{dx}E\hat{v}(\tilde{x}) < 0$. It can be checked algebraically that

$$\frac{d}{dx}E\hat{v}\left(\frac{2}{3}J + \frac{1}{3}\sqrt{J^2 - 3\hat{u}}\right) = -\frac{2}{27}\left[\left(J^2 - 3\hat{u}\right)^{\frac{3}{2}} - J(9\hat{u} + J^2)\right], \tag{40}$$

which is negative at least for all $J \in [0, 1]$ and $\hat{u} < 0$. Thus, x^*, which maximizes $E\hat{v}(x)$, is to the left of $\tilde{x}$. (Clearly, when $x^* \in (J, M)$ and $\tilde{x} = M$, it is trivial that x^* is to the left of $\tilde{x}$.)

3.2 "True override" regimes

Now, suppose we limit ourselves to reversion rules that merely involve the legislature's choosing a new point (and do not incorporate any other "penalties" for the judge). This seems reasonable: In the first place, the no-override range is only significant because the legislature will override the judge's decision when the decision is outside of the range; and in the second place, it is unclear whether the legislature can bring any other meaningful sanctions to bear on the judge.

Without reading the statute In case of an override, the legislature supersedes the point x that the judge established, instead establishing points $\hat{x}^C(x; C, P)$ or $\hat{x}^P(x; C, P)$ as described in section 2.4.

A bit of notation first: If the judge chooses x, the ultimate point that prevails is $y(x; C, P)$, where

$$y(x; C, P) \equiv \begin{cases} x \text{ if } x \in [C, P], \\ \hat{x}^C(x; C, P) \text{ if } x < C, \\ \hat{x}^P(x; C, P) \text{ if } x > P, \end{cases} \tag{41}$$

and so the judge's utility is $v(x) \equiv u(y(x; C, P))$.

The judge's expected utility (taking the expectation over C and P, though I will henceforth omit those subscripts on the expectation) is:

$$\begin{aligned} E_{C,P}\ v(x) \quad \equiv \quad & 2F(x)(1 - F(x))u(x) \\ & + \int_x^\infty \int_c^\infty u(\hat{x}^C(x; c, p))f_{C,P}(c, p)dpdc \\ & + \int_{-\infty}^x \int_{-\infty}^p u(\hat{x}^P(x; c, p))f_{C,P}(c, p)dcdp. \end{aligned} \tag{42}$$

To be more specific about the judge's utility, of course, we must commit ourselves to some specific form of override. I will discuss two different models of legislative overrides—the "closest point" rule and the "Nash bargaining" rule.

After reading the statute After reading the statute, the judge's expected utility is:

$$Ev(x) \equiv \frac{F(x)}{F(M)}u(x) + \int_x^M \int_M^\infty u(\hat{x}^C(x;c,p)) \frac{f(c)f(p)}{F(M)(1-F(M))} dpdc. \quad (43)$$

3.2.1 "Closest point" reversion rule

This subsection shows that not every statutory reversion rule makes the judge deviate from his ideal point J.

Suppose, for instance, the legislature, in case of override, chooses the "closest point" reversion rule, as described in section 2.4:

$$\begin{aligned} \hat{x}^C(x;C,P) &\equiv C, \\ \hat{x}^P(x;C,P) &\equiv P. \end{aligned} \quad (44)$$

Proposition 3 *Under the "closest point" reversion rule, the judge can do no better than to always choose his ideal point, $x = J$, regardless whether he reads the statute.*

Proof. (a) First, consider the case where the judge doesn't read the statute. Suppose the judge is considering ruling at any alternate point x, and note that $[C,P]$ is the no-override range. Then one of the following cases must hold:

- $J \in [C,P]$. If the judge chooses J, then J becomes the law. Then the judge does not benefit from choosing any other x—and in fact loses in all cases, except if J happens to be either C or P, x is outside of the $[C,P]$ range, and x is replaced by J.

- $J \notin [C,P]$ and $x \notin [C,P]$, and both J and x are on the same side of the override range. Then both J and x are overridden and replaced by the same closest point. J and x are equivalent to the judge, so the judge can do no better than to choose J.

- $J \notin [C,P]$ and $x \notin [C,P]$, but both J and x are on opposite sides of the override range. Then both J and x are overridden, but J is replaced by one endpoint of the range while x is replaced by another endpoint. Since J's replacement is the "closest point" of the no-override range to J, it is as least as good for the judge as x's replacement (and is strictly better if the range has positive length), and so the judge can do no better than to choose J.

- $J \notin [C,P]$ and $x \in [C,P]$. J is replaced by either C or P, which is at least as close to J as x (and is strictly closer unless x is at that same endpoint). Thus, the judge can do no better than to choose J.

The "closest point" reversion rule is thus "insufficiently punitive" to make the judge do anything but choose his own ideal point.

(b) Second, consider the case where the judge does read the statute. The judge will limit himself to the range $x \in [J, M]$, since a ruling at $x = J$ is better for the judge than any $x < J$, and a ruling at $x = M$ is better for the judge than any $x > M$.

After he reads the statute, the judge continues to choose his ideal point J. The proof is now even simpler. For any x between J and M, we have one of the following cases:

- $C \leq J$, that is, the no-override range contains J. The judge can do no better than to choose J.
- $C \in (J, x]$, that is, J is overridden by C, but x is not overridden. The judge does better by choosing J than by choosing x, because he prefers C to x.
- $C \in (x, M]$, that is, both J and x are overridden and replaced by the same point C. The judge is thus indifferent between J and x, and in particular can do no better than to choose J.

Just as before reading the statute, the "closest point" reversion rule is insufficiently punitive to make the judge do anything but choose his own ideal point. Whether the judge reads the statute or not makes no difference. ■

I omit an example in this case, since explaining it would be more complicated than the preceding proof.

3.2.2 "Nash bargaining" reversion rule

A more punitive reversion rule—and one that is probably more realistic—would be the "Nash bargaining" rule. Suppose, for instance, that the judge chooses some point $x < C$, and suppose that legislators have symmetric preferences around their ideal points.

Then the Congressman's best reversion point would be $\hat{x} = C$, though he would also settle for any point closer to C than x is. In other words, since x is to the left of C by a distance $C - x$, then the Congressman would be indifferent between x and a point that is $C - x$ to the *right* of C, that is, $C + C - x \equiv 2C - x$. Thus, the Congressman would be made better off by any point in the interval $(x, 2C - x)$, with C being his ideal. However, the Congressman and the President would unanimously agree so a movement as far as C, so after the unanimous possibilities are exhausted, the Congressman's negotiating range is $(C, 2C - x)$.

And, by the same logic, the President would be made better off by any point in $(x, 2P - x)$, with P being his ideal. However, both the Congressman and the President would want to move at least as far right as C, and neither would want to move to the right of P, so the President's negotiating range is (C, P).

Nash bargaining predicts that the two parties would settle on the midpoint of the overlap of their negotiating ranges. This overlap interval is $(C, 2C - x)$ if $2C - x < P$, and (C, P) otherwise. We thus have:

$$\hat{x}^C(x; C, P) \equiv \begin{cases} \frac{3C-x}{2} & \text{if } 2C - x < P, \\ \frac{C+P}{2} & \text{otherwise.} \end{cases} \tag{45}$$

Similarly, if $x > P$, then the President's negotiating range is $(2P - x, P)$ and the Congressman's negotiating range is (C, P). The overlap interval is $(2P - x, P)$ if $2P - x > C$, and (C, P) otherwise. We thus have:

$$\hat{x}^P(x; C, P) \equiv \begin{cases} \frac{3P-x}{2} & \text{if } 2P - x > C, \\ \frac{C+P}{2} & \text{otherwise.} \end{cases} \tag{46}$$

Without reading the statute Then (42) becomes:

$$\begin{aligned} Ev_N(x) \equiv\ & 2F(x)(1 - F(x))u(x) \\ & + \int_x^{\infty} \int_c^{2c-x} u\left(\frac{c+p}{2}\right) f_{C,P}(c,p)dpdc \\ & + \int_x^{\infty} \int_{2c-x}^{\infty} u\left(\frac{3c-x}{2}\right) f_{C,P}(c,p)dpdc \\ & + \int_{-\infty}^{x} \int_{-\infty}^{2p-x} u\left(\frac{3p-x}{2}\right) f_{C,P}(c,p)dcdp \\ & + \int_{-\infty}^{x} \int_{2p-x}^{p} u\left(\frac{c+p}{2}\right) f_{C,P}(c,p)dcdp, \end{aligned} \tag{47}$$

and the first-order condition is (after several cancellations):

$$\begin{aligned} \frac{d}{dx} Ev_N(x^*) \equiv\ & 2F(x^*)(1 - F(x^*))u'(x^*) + 2f(x^*)(1 - 2F(x^*))u(x^*) \\ & - \int_{x^*}^{\infty} u(x^*) f_{C,P}(x^*, p)dp + \int_{-\infty}^{x^*} u(x^*) f_{C,P}(c, x^*)dc \\ & - \frac{1}{2} \int_{x^*}^{\infty} \int_{2c-x^*}^{\infty} u'\left(\frac{3c-x^*}{2}\right) f_{C,P}(c,p)dpdc \\ & - \frac{1}{2} \int_{-\infty}^{x^*} \int_{-\infty}^{2p-x^*} u'\left(\frac{3p-x^*}{2}\right) f_{C,P}(c,p)dpdc \\ =\ & 0. \end{aligned} \tag{48}$$

This, in turn, can be simplified to:

$$\begin{aligned} \frac{d}{dx} Ev_N(x^*) \equiv\ & 2F(x^*)(1 - F(x^*))u'(x^*) \\ & - \int_{x^*}^{\infty} u'\left(\frac{3c-x^*}{2}\right) f(c)(1 - F(2c - x^*))dc \\ & - \int_{-\infty}^{x^*} u'\left(\frac{3p-x^*}{2}\right) f(p)F(2p - x^*)dp \\ =\ & 0. \end{aligned} \tag{49}$$

We can see that $x^* = J$ is a solution if $J = L_M$ (that is, if the judge's ideal point happens to be the median of the distribution of legislators' positions), and if we also assume that $f(x)$ is symmetric around L_M and that $u(x)$ is symmetric around the judge's ideal point J. To see this, plug $x^* = J$ into (49). The first term vanishes because $u'(J) = 0$. As for the second and third terms, they cancel each other out; one can derive this by observing, first, that $u'(x) = -u'(2J - x)$ (from the symmetry of u around J); second, that $f(x) = f(2J - x)$ (from the symmetry of f around $L_M = J$); third, that $1 - F(2C - J) = F(3J - 2C)$ (this follows from the second point); and, finally, by introducing a change of variables $t = 2J - C$. It is intuitively clear that $x^* = J$ is not just a critical point of $\frac{d}{dx}Ev_N(x)$, but also a minimum.

Because this expression is complicated, it will be more convenient to continue the illustration with an example.

Example 4

Consider again the example where L is uniformly distributed on $[0, 1]$, and $u(x) \equiv -(x - J)^2$. We must alter the calculation of the judge's expected utility to take into account the fact that policy points are limited to the finite interval $[0, 1]$. The reversion points are:

$$\hat{x}^C(x; C, P) \equiv \begin{cases} \frac{3C-x}{2} \text{ if } 2C - x < 1 \text{ and } 2C - x < P, \\ \frac{C+P}{2} \text{ if } 2C - x > 1 \text{ or } 2C - x > P, \end{cases} \tag{50}$$

and:

$$\hat{x}^P(x; C, P) \equiv \begin{cases} \frac{3P-x}{2} \text{ if } 2P - x > 0 \text{ and } 2P - x > C, \\ \frac{C+P}{2} \text{ if } 2P - x < 0 \text{ or } 2P - x < C. \end{cases} \tag{51}$$

The judge's expected utility is thus:

$$\begin{aligned} Ev_N(x) \quad \equiv \quad & 2F(x)(1 - F(x))u(x) \\ & + \int_x^{\frac{x+1}{2}} \int_c^{2c-x} u\left(\frac{c+p}{2}\right) f_{C,P}(c,p)dpdc \\ & + \int_x^{\frac{x+1}{2}} \int_{2c-x}^{1} u\left(\frac{3c-x}{2}\right) f_{C,P}(c,p)dpdc \\ & + \int_{\frac{x+1}{2}}^{1} \int_c^{1} u\left(\frac{c+p}{2}\right) f_{C,P}(c,p)dpdc \\ & + \int_0^{\frac{x}{2}} \int_0^{p} u\left(\frac{c+p}{2}\right) f_{C,P}(c,p)dcdp \\ & + \int_{\frac{x}{2}}^{x} \int_0^{2p-x} u\left(\frac{3p-x}{2}\right) f_{C,P}(c,p)dcdp \\ & + \int_{\frac{x}{2}}^{x} \int_{2p-x}^{p} u\left(\frac{c+p}{2}\right) f_{C,P}(c,p)dcdp, \end{aligned} \tag{52}$$

and the first-order condition is (after several cancellations):

$$\begin{aligned}\frac{d}{dx}Ev_N(x^*) \equiv\;& 2F(x^*)(1-F(x^*))u'(x^*)+2f(x^*)(1-2F(x^*))u(x^*)\\ &-\int_{x^*}^{1}u(x^*)f_{C,P}(x^*,p)dp+\int_0^{x^*}u(x^*)f_{C,P}(c,x^*)dc\\ &-\frac{1}{2}\int_{x^*}^{\frac{x^*+1}{2}}\int_{2c-x^*}^{1}u'\left(\frac{3c-x^*}{2}\right)f_{C,P}(c,p)dpdc\\ &-\frac{1}{2}\int_{\frac{x^*}{2}}^{x^*}\int_0^{2p-x^*}u'\left(\frac{3p-x^*}{2}\right)f_{C,P}(c,p)dpdc\\ =\;& 0. \end{aligned} \tag{53}$$

This, in turn, can be simplified to:

$$\frac{d}{dx}Ev_N(x^*)\equiv\frac{19}{4}x^{*3}-x^{*2}\left(5J+\frac{37}{8}\right)+x^*\left(5J+\frac{1}{8}\right)-\frac{1}{2}J+\frac{1}{8}=0. \tag{54}$$

We can calculate that $\frac{d}{dx}Ev_N(J)=-\frac{1}{4}J^3+\frac{3}{8}J^2-\frac{3}{8}J+\frac{1}{8}$, which is equal to 0 if $J=\frac{1}{2}$, and the second derivative, $\frac{d^2}{dx^2}Ev_N(J)=\frac{17}{4}J^2-\frac{17}{4}J+\frac{1}{8}$, is negative at $J=\frac{1}{2}$; so it is optimal for the judge to stay at his ideal point if his ideal point is the median.

When $J\neq\frac{1}{2}$, we can also verify that $\frac{d}{dx}Ev_N(J)$ and $\frac{d}{dx}Ev_N\left(\frac{1}{2}\right)$ are of different signs: $\frac{d}{dx}Ev_N(J)=-\frac{1}{4}J^3+\frac{3}{8}J^2-\frac{3}{8}J+\frac{1}{8}$ and $\frac{d}{dx}Ev_N\left(\frac{1}{2}\right)=\frac{3}{4}J-\frac{3}{8}$, and the product of these is strictly negative when $J\neq\frac{1}{2}$. Therefore, $\frac{d}{dx}Ev_N(x)$ must have a zero between J and $\frac{1}{2}$.

After reading the statute If $C\in\left[x,\frac{x+M}{2}\right]$, then the Congressman would agree to any reversion in the range $[x,2C-x]$; and since $2C-x<M$, it is clear that P is outside of this range. Thus, Nash bargaining will lead to the reversion point $\frac{3C-x}{2}$.

If $C\in\left[\frac{x+M}{2},M\right]$, then we must distinguish two cases. The Congressman would still agree to any reversion in the range $[x,2C-x]$, so Nash bargaining will lead to the reversion point $\frac{3C-x}{2}$ as long as $P>2C-x$. However, if $P<2C-x$, then Nash bargaining leads to the reversion point $\frac{C+P}{2}$.

Thus, the reversion points will be the following:

$$\hat{x}^C(x;C,P)\equiv\begin{cases}\frac{3C-x}{2} & \text{if } C<\frac{x+M}{2},\\ \frac{3C-x}{2} & \text{if } C>\frac{x+M}{2} \text{ and } P>2C-x,\\ \frac{C+P}{2} & \text{if } C>\frac{x+M}{2} \text{ and } P<2C-x.\end{cases} \tag{55}$$

The judge's expected utility will then be (referring back to (43)):

$$\begin{aligned} Ev_N^M(x) \quad &\equiv \quad \frac{F(x)}{F(M)}u(x) + \int_x^{\frac{x+M}{2}} u\left(\frac{3c-x}{2}\right)\frac{f(c)}{F(M)}dc \\ &\quad + \frac{1}{F(M)(1-F(M))}\int_{\frac{x+M}{2}}^{M}\left[\begin{array}{c} \int_M^{2c-x} u\left(\frac{c+p}{2}\right)f(p)dp \\ +u\left(\frac{3c-x}{2}\right)\int_{2c-x}^{\infty} f(p)dp \end{array}\right] f(c)dc \\ &\equiv \quad \frac{F(x)}{F(M)}u(x) + \int_x^{\frac{x+M}{2}} u\left(\frac{3c-x}{2}\right)\frac{f(c)}{F(M)}dc \\ &\quad + \frac{1}{F(M)(1-F(M))}\int_{\frac{x+M}{2}}^{M}\left[\begin{array}{c} \int_M^{2c-x} u\left(\frac{c+p}{2}\right)f(p)dp \\ +u\left(\frac{3c-x}{2}\right)(1-F(2c-x)) \end{array}\right] f(c)dc. \end{aligned} \tag{56}$$

I will continue with an example.

Example 5

Consider again the example where L is uniformly distributed on $[0, 1]$, and $u(x) \equiv -(x-J)^2$. We again alter the reversion points to take account of the boundaries of the policy spectrum $[0, 1]$. If $\frac{x+1}{2} < M$ (that is, if $x < 2M - 1$), the reversion points are:

$$\hat{x}^C(x; C, P) \equiv \begin{cases} \frac{3C-x}{2} \text{ if } C < \frac{x+M}{2}, \\ \frac{C+P}{2} \text{ if } C \in \left(\frac{x+M}{2}, \frac{x+1}{2}\right) \text{ and } P < 2C - x, \\ \frac{3C-x}{2} \text{ if } C \in \left(\frac{x+M}{2}, \frac{x+1}{2}\right) \text{ and } P > 2C - x, \\ \frac{C+P}{2} \text{ if } C > \frac{x+1}{2}, \end{cases} \tag{57}$$

and so the judge's expected utility is:

$$\begin{aligned} Ev_N^M(x) \quad &\equiv \quad -\frac{x(x-J)^2}{M} - \frac{1}{M}\int_x^{\frac{x+M}{2}}\left(\frac{3c-x}{2} - J\right)^2 dc \\ &\quad -\frac{1}{M(1-M)}\int_{\frac{x+M}{2}}^{\frac{x+1}{2}}\left[\begin{array}{c} \int_M^{2c-x}\left(\frac{c+p}{2} - J\right)^2 dp \\ +\left(\frac{3c-x}{2} - J\right)^2(1-2c+x) \end{array}\right] dc \\ &\quad -\frac{1}{M(1-M)}\int_{\frac{x+1}{2}}^{M}\int_M^1\left(\frac{c+p}{2} - J\right)^2 dpdc. \end{aligned} \tag{58}$$

And if $\frac{x+1}{2} > M$ (that is, if $x > 2M - 1$), the reversion points are:

$$\hat{x}^C(x; C, P) \equiv \begin{cases} \frac{3C-x}{2} \text{ if } C < \frac{x+M}{2}, \\ \frac{C+P}{2} \text{ if } C > \frac{x+M}{2} \text{ and } P < 2C - x, \\ \frac{3C-x}{2} \text{ if } C > \frac{x+M}{2} \text{ and } P > 2C - x, \end{cases} \tag{59}$$

and so the judge's expected utility is:

$$\begin{aligned} Ev_N^M(x) \quad &\equiv \quad -\frac{x(x-J)^2}{M} - \frac{1}{M}\int_x^{\frac{x+M}{2}}\left(\frac{3c-x}{2} - J\right)^2 dc \\ &\quad -\frac{1}{M(1-M)}\int_{\frac{x+M}{2}}^{M}\left[\begin{array}{c} \int_M^{2c-x}\left(\frac{c+p}{2} - J\right)^2 dp \\ +\left(\frac{3c-x}{2} - J\right)^2(1-2c+x) \end{array}\right] dc. \end{aligned} \tag{60}$$

To keep this example manageable, let us consider some numerical examples.

Consider $J = 0.1$. By (54), if the judge does not read the statute, he rules at approximately 0.235. The median of the legislators' distribution (0.5) exerts some pull on him, because he does not want to be overridden. Now consider three different examples of what the judge could discover by reading the statute:

- First, suppose he reads the statute and finds that $M = 0.5$. He maximizes his expected utility as expressed in (60) (since $2M - 1 = 0$). He now rules at approximately 0.245. Reading the statute has made him deviate *toward* the statute.
- Second, suppose he reads the statute and finds that $M = 0.4$. Maximizing the same portion of the expected utility function (since $2M - 1 = -0.2 < 0$), he now rules at approximately 0.206. Reading the statute has made him deviate *away* from the statute and back toward his ideal point. Reading the statute has given him good news—he has discovered that the no-override range was closer to his ideal point than he had feared. Instead of deviating toward the median of the ex-ante distribution (0.5), he could now deviate toward the statutory point, which is closer to him. Thus, he can now afford to rule closer to his ideal point.
- Third—just to illustrate the interaction of (58) and (60)—suppose he reads the statute and finds that $M = 0.9$. Since $2M - 1 = 0.8$, he must focus on two ranges—$x \in [0.1, 0.8]$ (for which he maximizes his expected utility as expressed in (he maximizes his expected utility as expressed in (58)) and $x \in [0.8, 0.9]$ (for which he maximizes his expected utility as expressed in (60)). Maximizing (60), we find that there is no interior minimum; therefore, we check the value of expected utility at the bounds of the domain and obtain -0.502 for $x = 0.8$ and -0.64 for $x = 0.9$. Now focusing on the range $x \in [0.1, 0.8]$, we maximize (58) and obtain an interior minimum of $x = 0.323$, which gives an expected utility of -0.266. Therefore, the judge rules at 0.323, meaning that he moves *toward* the statute.

3.3 Discussion

The previous sections have shown one very intuitive result—that the "very large penalty" rule makes the judge deviate toward the median of the legislative distribution, and that reading the statute makes him deviate again toward the statute. They have also shown a few surprising results.

Ineffectiveness of the "closest point" reversion rule First, under the "closest point" reversion rule, the judge never deviates from his ideal point, whether or not he reads the statute.

Basically, this rule is not punitive enough. The rule may *seem* punitive—if the judge rules at $x = 0.2$ and the override range is $[0.4, 0.7]$, then the override will be to 0.4, which does seem bad. But it is not punitive *on the margin.*

Suppose the judge starts out at some x between his ideal point J and the median L_M, or between his ideal point and the statutory point M; and suppose he is considering a small movement toward J. He experiences a first-order increase in his utility from moving closer to his ideal point—roughly speaking, the probability of no-override times $u'(x)$. However, the decrease in his utility from the increase in override probability is only second-order, because if this small movement causes an override, this means that he has just crossed the position of a pivotal legislator, say $x = C$, and under the "closest point" rule, the override point is just equal to C.

Therefore, nothing prevents him from moving all the way to J. So a reversion rule will only affect the judge if it is sufficiently punitive on the margin.

Complex effects of the "Nash bargaining" reversion rule Second, under the "Nash bargaining" rule, the judge is not guaranteed to move closer to the statutory point M, relative to where he would have moved if he had not read the statute. Reading a statute may have effects that pull in different directions.

First, reading a statute gives the judge better information about where legislative preferences lie, so his previous optimum will in general no longer be optimal. Because M is a safe haven—he will never be overruled at M provided legislative preferences have not changed—he may seek to avoid override by going toward M (up to a possible corner solution at M).

However, reading a statute may also give the judge good news about legislative preferences; to the extent he was seeking to avoid an override by moving toward the median, he may now decide that he can afford to move closer to his ideal point J.

Thus, the existence of statutes has a complex and unpredictable effect on judges' compliance with the law. We cannot assume that the existence of a statute increases the likelihood that the judge will rule "according to the statute" (if we define "according to the statute" as a continuous concept meaning "closer to M").

A concluding note on statutory interpretation I have called this exercise "statutory interpretation," with the statute establishing a "most plausible point." However, the judge only consults the statute to find out "which way the wind is blowing," i.e., what he can get away with without being overridden.

This seems to conflict with the practice of judicial opinion writing. A judge not only rules a certain way but also tries to justify his ruling as a fair (or even the best) interpretation of the statute. What does it mean to say, within this model, that a judge with $J = 0.1$ reads a statute with $M = 0.4$ and rules at $\tilde{x} = 0.25$?

The answer is that this model implicitly accepts the "indeterminacy" critique of Legal Realists and Critical Legal Scholars that holds that legal materials do not significantly constrain the judge. The judge decides how to rule on ideological grounds alone—maximizing his ideological utility, taking into account the possibility of legislative overrides—and the language justifying the ruling as an interpretation of the statute is mere rhetoric.

4 Adding a second method of interpretation

4.1 Setup

Now suppose that, instead of a single method of statutory interpretation, there are two available methods of statutory interpetation, for instance textualism and intentionalism. (I have defined these terms in the Introduction.) In other words, suppose that, consulting standard textual sources (like dictionaries or examples of usage contemporaneous with the enactment of the statute) the most plausible point is M_1, while consulting standard intentionalist sources (like records of legislative debate or committee reports) the most plausible point is M_2. (Assume, without loss of generality, that $M_1 < M_2$.)

Can interpretive methods make a difference? Usually most methods of statutory interpretation lead to a common conclusion, and indeed, the previous assumption that there is only one available method can be alternatively construed as an assumption that all available methods lead to a common point M.

In Volokh (2008, pp. 779–80), I provide several examples of statutes where the interpretive method does seem to make a difference:

- The Individuals with Disabilities Education Act allows prevailing parents to recover "reasonable attorneys' fees as part of the costs." Does this allow the recovery of *expert witness* fees? The Second Circuit and the Supreme Court both agreed that the statutory text did not allow the recovery of such fees, but that legislative history pointed the other way. (See *Murphy v. Arlington Cent. Sch. Dist. Bd. of Educ.*, 402 F.3d 332 (2d Cir. 2005); *Arlington Cent. Sch. Dist. Bd. of Educ. v. Murphy*, 548 U.S. 291 (2006).)

- The Clean Water Act defines "pollutant" to include "radioactive materials." Does this cover radioactive materials that are already regulated by the Atomic Energy Commission? The Tenth Circuit and the Supreme Court both agreed that they were covered under the statutory text, but that the legislative history "speaks with force" in the other direction. (See *Colo. Pub. Int. Res. Group v. Train*, 507 F.2d 743 (10th Cir. 1974); *Train v. Colo. Pub. Int. Res. Group*, 426 U.S. 1 (1976).)

- The act creating Chapter 12 of the Bankruptcy Code, which applies to family farmers, provided that the provisions of Chapter 12 did not apply to bankruptcy cases "commenced . . . before the effective date of this Act." However, the legislative history said the opposite—that cases already pending could be converted to Chapter 12 in the "sound discretion" of courts, "where it is equitable to do so." (See *In re Sinclair*, 870 F.2d 1340 (7th Cir. 1989).)

More commonly, different interpretive methods do not *mandate* one result or another, but they do make one result or another *more plausible* because of the different sources they rely on. For instance, one can imagine a statute where two results are possible, and either result could be argued—with enough ingenuity—using either method, but the sources used under one method point more strongly to one result while the sources used under another method point more strongly to the other result.

Or, even less determinately, we can imagine that *every* result is possible, but that, if a judge wants to argue for a particular result, he may choose to rely on the reasoning of one method rather than the other. (As in the previous Part, the method does not drive the result, but is a rhetoric used to rationalize the result that the judge has reached by other means.)

Why would this happen? Why would a legislature act in this way? If a judge observes a statute with two distinct most plausible points, M_1 and M_2, depending on what method he uses, what should he infer about the locations of C and P?

There may be strategic reasons why legislators may act in this way, but here, I take a simpler path and assume that the existence of two points is inadvertent. The legislators may have been trying to implement M_1 but accidentally gave signals that the meaning of the statute is M_2, or vice versa. If the judge knows whether the legislators were shooting for M_1 or M_2, he may safely ignore the "accidental point," and we are back in the previous case. However, now let us suppose that the judge does not know what point the legislators intended. Instead, he believes that the true point is M_1 with probability q and M_2 with probability $1 - q$.

The legislators' new distributions Before reading the statute, the density functions of C and P are as described above. After reading the statute, they become more complicated. To begin with, when there was only a point M, we could conclude that $C < M < P$, so $\phi_C(x) \equiv 0$ for $x > M$ and $\phi_P(x) \equiv 0$ for $x < M$. There was never any interval where both $\phi_C(x)$ and $\phi_P(x)$ could be non-zero. However, with two points M_1 and M_2, we know that $C < M_2$ and that $P > M_1$, so both $\phi_C(x)$ and $\phi_P(x)$ are potentially non-zero in the interval $[M_1, M_2]$.

After reading the statute, the distributions of C and P depend on the interval. When $x < M_1$:

$$\begin{cases} \phi_C(x) \equiv q\frac{f(x)}{F(M_1)} + (1-q)\frac{f(x)}{F(M_2)}, \\ \phi_P(x) \equiv 0; \end{cases} \tag{61}$$

when $x \in (M_1, M_2)$:

$$\begin{cases} \phi_C(x) \equiv (1-q)\frac{f(x)}{F(M_2)}, \\ \phi_P(x) \equiv q\frac{f(x)}{1-F(M_1)}; \end{cases} \tag{62}$$

and when $x > M_2$:

$$\begin{cases} \phi_C(x) \equiv 0, \\ \phi_P(x) \equiv q\frac{f(x)}{1-F(M_1)} + (1-q)\frac{f(x)}{1-F(M_2)}. \end{cases} \tag{63}$$

4.2 Judicial behavior under the "very large penalty" model

For simplicity, let us consider the "very large penalty" model. And without loss of generality, let us restrict ourselves to the case where $x < M_2$. We must therefore distinguish two cases: $x < M_1$ and $x \in (M_1, M_2)$ (though, for the sake of completeness, I will define $Ev^M(x)$ for $x > M_2$ as well).

Case 1: The left side First, suppose $x < M_1$. We have:

$$\begin{aligned} Ev^M(x) &\equiv \left(q\frac{F(x)}{F(M_1)} + (1-q)\frac{F(x)}{F(M_2)}\right)u(x) \\ &\quad + \left(1 - \left(q\frac{F(x)}{F(M_1)} + (1-q)\frac{F(x)}{F(M_2)}\right)\right)\hat{u} \\ &\equiv \left(q\frac{1}{F(M_1)} + (1-q)\frac{1}{F(M_2)}\right)F(x)(u(x)-\hat{u}) + \hat{u} \\ &\equiv \frac{qF(M_2) + (1-q)F(M_1)}{F(M_1)F(M_2)}F(x)(u(x)-\hat{u}) + \hat{u}, \end{aligned} \tag{64}$$

and so:

$$\frac{d}{dx}Ev^M(x) \equiv \frac{qF(M_2) + (1-q)F(M_1)}{F(M_1)F(M_2)}[f(x)(u(x)-\hat{u}) + F(x)u'(x)]. \tag{65}$$

Note that within this range, $\frac{d}{dx}Ev^M(x) > 0$ for $\hat{u}$ sufficiently low or for $u'(x)$ near zero.

Case 2: The middle range Second, suppose $x \in (M_1, M_2)$. Now we have:

$$\begin{aligned} Ev^M(x) &\equiv \left(q\frac{1-F(x)}{1-F(M_1)} + (1-q)\frac{F(x)}{F(M_2)}\right)u(x) \\ &\quad + \left(1 - \left(q\frac{1-F(x)}{1-F(M_1)} + (1-q)\frac{F(x)}{F(M_2)}\right)\right)\hat{u} \\ &\equiv \left(q\frac{1-F(x)}{1-F(M_1)} + (1-q)\frac{F(x)}{F(M_2)}\right)(u(x)-\hat{u}) + \hat{u} \\ &\equiv \frac{qF(M_2)(1-F(x)) + (1-q)(1-F(M_1))F(x)}{(1-F(M_1))F(M_2)}(u(x)-\hat{u}) + \hat{u} \\ &\equiv \frac{qF(M_2) + F(x)((1-q)(1-F(M_1)) - qF(M_2))}{(1-F(M_1))F(M_2)}(u(x)-\hat{u}) + \hat{u} \\ &\equiv \frac{\alpha + \beta F(x)}{\gamma}(u(x)-\hat{u}) + \hat{u}, \end{aligned} \tag{66}$$

where $\alpha \equiv qF(M_2)$, $\beta \equiv (1-q)(1-F(M_1))-qF(M_2)$, and $\gamma \equiv (1-F(M_1))F(M_2)$. Thus:

$$\frac{d}{dx}Ev^M(x) \equiv \frac{\beta}{\gamma}f(x)(u(x)-\hat{u}) + \frac{\alpha+\beta F(x)}{\gamma}u'(x). \tag{67}$$

Note that within this range, for $\hat{u}$ sufficiently low or for $u'(x)$ near zero, $\frac{d}{dx}Ev^M(x)$ has the same sign as β.

Case 3: The right side Third, for completeness (this case is not relevant given our assumption that $J < M_2$), suppose $x > M_2$. We have:

$$\begin{aligned} Ev^M(x) &\equiv \left(q\frac{1-F(x)}{1-F(M_1)} + (1-q)\frac{1-F(x)}{1-F(M_2)}\right)u(x) \\ &\quad + \left(1 - \left(q\frac{1-F(x)}{1-F(M_1)} + (1-q)\frac{1-F(x)}{1-F(M_2)}\right)\right)\hat{u} \\ &\equiv \left(q\frac{1}{1-F(M_1)} + (1-q)\frac{1}{1-F(M_2)}\right)(1-F(x))(u(x)-\hat{u}) + \hat{u} \\ &\equiv \frac{q(1-F(M_2)) + (1-q)(1-F(M_1))}{(1-F(M_1))(1-F(M_2))}(1-F(x))(u(x)-\hat{u}) + \hat{u} \\ &\equiv \frac{1-qF(M_2)) - (1-q)F(M_1)}{(1-F(M_1))(1-F(M_2))}(1-F(x))(u(x)-\hat{u}) + \hat{u}, \end{aligned} \tag{68}$$

and so:

$$\frac{d}{dx}Ev^M(x) \equiv \frac{1-qF(M_2)+(1-q)F(M_1)}{F(M_1)F(M_2)}[-f(x)(u(x)-\hat{u}) + (1-F(x))u'(x)]. \tag{69}$$

Note that within this range, $\frac{d}{dx}Ev^M(x) < 0$ for $\hat{u}$ sufficiently low or for $u'(x)$ near zero.

Results It is easier to describe what happens if $J \in (M_1, M_2)$, so let us consider this case first.

It is clear that the judge will rule at some point within (M_1, M_2) if his ideal point is within that range, as he can never do better by ruling outside of that range. But he may move left or right within that range, depending on the sign of the derivative in (67). Since $u'(J) = 0$, which direction the judge moves depends on the sign of $\beta \equiv (1-q)(1-F(M_1)) - qF(M_2)$, that is, on whether $F(M_1) + \frac{q}{1-q}F(M_2) < 1$.

- If $\beta < 0$, then $\frac{d}{dx}Ev^M(x) < 0$, and so the judge will move left, in the direction of M_1. This will be the case, for instance, if "most judges are on the left," as $F(M_1)$ and $F(M_2)$ will then be high. If $\frac{d}{dx}Ev^M(x)$ continues to be negative over the whole range (M_1, J), the judge will reach a corner solution at M_1; but this does not necessarily happen.

- Conversely, if $\beta > 0$, then $\frac{d}{dx}Ev^M(x) > 0$, and so the judge will move right, in the direction of M_2. This will be the case, for instance, if "most judges are on the right," as $F(M_1)$ and $F(M_2)$ will then be low. If $\frac{d}{dx}Ev^M(x)$ continues to be positive over (J, M_2), the judge will reach a corner solution at M_2; but again, this does not necessarily happen.

Now let us consider what happens if $J < M_1$. Since $u'(J) = 0$, the derivative in (65) is positive, and so the judge will move right. If $\frac{d}{dx}Ev^M(x)$ continues to be negative over the whole range (J, M_1), he may move as far as M_1 itself, and there may be a corner solution at this point.

The judge may move even further right, in the direction of M_2 (and up to a possible corner solution at M_2), but this is not guaranteed, even if $\beta > 0$. For instance, suppose that, at M_1, the left-derivative of $Ev^M(x)$ (as given by (65)) is positive while the right-derivative of $Ev^M(x)$ (as given by (67)) is negative. That is:

$$\begin{cases} f(M_1)(u(M_1) - \hat{u}) + F(M_1)u'(M_1) > 0, \\ \frac{\beta}{\gamma}f(M_1)(u(M_1) - \hat{u}) + \frac{\alpha+\beta F(M_1)}{\gamma}u'(M_1) < 0. \end{cases} \tag{70}$$

But this implies (if $\beta > 0$):

$$\begin{cases} \beta f(M_1)(u(M_1) - \hat{u}) + \beta F(M_1)u'(M_1) > 0, \\ \beta f(M_1)(u(M_1) - \hat{u}) + (\alpha + \beta F(M_1))\, u'(M_1) < 0, \end{cases} \tag{71}$$

or, in other words, that $\alpha u'(M_1)$ is sufficiently negative—which is entirely possible if, say, J is far enough left that the judge's utility is already dropping precipitously by the time he rules at M_1.

Thus, different judges will move toward different points. If $\beta > 0$, judges within (M_1, M_2) will all move toward M_2. However, judges to the left of M_1 will move toward M_1, and some might stay there at a corner solution. The existence of two "most plausible points" of the statute, depending on the method of interpretation, gives rise to two "points of attraction" for judges, and this affects the observed distribution of judges.

I will illustrate this point with an example.

Example 6

Consider again the example where L is uniformly distributed on $[0, 1]$, $u(x) \equiv -(x - J)^2$, and $q = \frac{1}{2}$. Now suppose that $M_1 = 0.3$ and $M_2 = 0.6$. (This implies that $\beta > 0$.) Finally, suppose that $\hat{u} = -0.15$. Imagine 101 different judges whose ideal points J are uniformly distributed on $[0, 1]$, that is, 0, 0.01, 0.02, 0.03, and so on.

Using numerical methods, I maximized $Ev^M(x)$ for each of these judges under different assumptions.

All judges use textualism First, suppose textualism is the only available interpretive method, so there is only one statutory point, $M = 0.3$. Using the expressions for $Ev^M(x)$ in (30) and (32), we obtain the following:

- Judges whose ideal points lie in $[0, 0.05]$ move right, toward $M = 0.3$; their $\tilde{x}$ are all above 0.2.
- Judges whose ideal points lie in $[0.06, 0.40]$ reach a corner solution at $\tilde{x} = M = 0.3$.
- Judges whose ideal points lie in $[0.41, 1]$ move left, toward $M = 0.3$; their $\tilde{x}$ are all below 0.78.

Under a mandatory textualist regime, the distribution of judicial decisions—which is observable—thus has a very large mass point at $M = 0.3$ (with probability 0.35), and has a mean of $\bar{T} = 0.446$ with variance 0.028.

All judges use intentionalism Now, suppose intentionalism is the only available interpretive method, so there is only one statutory point, $M = 0.6$. In the same way as above, we obtain the following:

- Judges in $[0, 0.48]$ move right, toward $M = 0.6$; their $\tilde{x}$ are all above 0.2.
- Judges in $[0.49, 0.81]$ reach a corner solution at $\tilde{x} = M = 0.6$.
- Judges in $[0.82, 1]$ move left, toward $M = 0.6$; their $\tilde{x}$ are all below 0.78.

Under a mandatory intentionalist regime, the distribution of judicial decisions thus has a very large mass point at $M = 0.6$ (with probability 0.33), and has a mean of $\bar{I} = 0.525$ with variance 0.022.

A "choice regime" Finally, suppose we have a "choice regime," where two interpretive methods are available to judge, so M_1 and M_2 are 0.3 and 0.6, as given above. This is, by and large, the system we have today. Now some judges will move toward M_1 and others will move toward M_2.

- Judges in $[0, 0.11]$ move right, toward $M_1 = 0.3$.
- Judges in $[0.12, 0.28]$ hit a corner solution at $\tilde{x} = M_1 = 0.3$.
- Judges in $[0.29, 0.3]$ move *beyond* the corner solution and end up above $M_1 = 0.3$.
- Judges in $[0.31, 0.58]$ move right, toward $M_2 = 0.6$.
- Judges in $[0.59, 0.6]$ hit the corner solution at $\tilde{x} = M_2 = 0.6$.
- Judges in $[0.61, 0.75]$ likewise hit the corner solution at $\tilde{x} = M_2 = 0.6$, though from the other direction.
- Judges in $[0.76, 1]$ move left, toward $M_2 = 0.6$.

Under a choice regime, the distribution of judicial decisions thus has two mass points at $M_1 = 0.3$ and $M_2 = 0.6$ (each with probability 0.17), and has a mean of $\bar{C} = 0.486$ with variance 0.028.

Misleading conclusions one can draw from the choice regime Unsurprisingly, we see that when judges can choose their interpretive method—that is, when they can move toward either M_1 or M_2—the result is more "moderate" than when only one statutory point is available. In this example, the mean when there are two methods, $\bar{C} = 0.486$, is closer to the true mean of judicial ideology (0.5) than when only one method can be used (in which case we get $\bar{T}$ or $\bar{I}$).

However, we can observe more than just the judicial decision. Every opinion must rely on *some* interpretive method, because the judge will rely on some sources and not others; sometimes, the opinion relies on several methods at once. Let us suppose, then, that the judges who move toward M_1 use textualist reasoning in their decision, while judges who move toward M_2 use intentionalist reasoning, so that an outside observer could code the opinions as "textualist" or "intentionalist." (It may be more realistic to code them otherwise—for instance, perhaps judges who settle on $x < \frac{1}{2}(M_1 + M_2)$ should be coded as textualists while judges who settle on $x > \frac{1}{2}(M_1 + M_2)$ should be coded as intentionalists. The most important thing, though, is that somehow, all judges be coded as one or the other.)

Thus, the judges between 0 and 0.28, who moved toward, or hit a corner solution at, $M_1 = 0.3$, will be observed to write textualist opinions in a choice regime, while the judges between 0.29 and 1, who moved toward, or hit a corner solution at, $M_2 = 0.6$, will be observed to write intentionalist opinions. *Observed* textualist decisions have a mean of $\bar{T}_O = 0.284$ with variance 0.0006, while *observed* intentionalist decisions have a mean of $\bar{I}_O = 0.568$ with variance 0.016.

Thus, suppose we are trying to evaluate the common claim that textualism is a "conservative" method. (Textualism has been called "neo-conservative," or "antigovernmental," or characterized as antiregulatory, see Volokh (2008, pp. 771–72); Eskridge & Frickey (1994, p. 77); Marmor (2005, pp. 2064 & n.3, 2066); Mank (1996, p. 1233); Eskridge (1991, 410); Easterbrook (1988, p. 65).) This claim is usually supported by reference to observed textualist opinions. In this example, the mean of observed textualist opinions is $\bar{T}_O = 0.284$, which is more extreme than the mean of either the legislative or the judicial distribution (both assumed to be 0.5 here), and significantly different than the mean of observed intentionalist opinions, which is $\bar{I}_O = 0.568$. (Note that "left" and "right" do not have their common political meanings in this example.)

But such an analysis would substantially overstate the "true effect" of textualism or intentionalism. The mean of observed textualist opinions, $\bar{T}_O = 0.284$, is in this case more extreme than textualism's statutory point, $M_1 = 0.3$. (This is not true in general; observe that $\bar{I}_O$ is less extreme than M_2.) More importantly, it is also substantially more extreme than the mean that would result if textualism were forced on the entire judiciary, for instance if there were Federal Rules of Statutory Interpretation (see Rosenkranz (2002)). The mean of textualism in that case would $\bar{T} = 0.446$—not too far off from the mean of intentionalism if mandated for everyone, which is $\bar{I} = 0.525$. Relying on observed opinions using a particular method may tell us more about the judges who choose to use the method than about the method itself. It may tell us more about *textualists* than about *textualism*.

In other words, inferring the political bias of an interpretive method from observing published decisions can substantially overstate the "true" bias of the method. Empirical strategies that have been used to discuss interpretive methods must be rethought to take the self-selection aspect into account.

5 Conclusion

This paper shows two sets of results, one when there is only interpretive method and another when there are several methods available.

One interpretive method When a single method of interpretation is available, the judge will often benefit from reading the statute when the legislature can penalize him. If the judge knows the location of the "no-override range," he doesn't need to read the statute, but reading the statute may benefit him when is uncertain of legislative preferences. For instance, when the legislature can impose a simple penalty on the judge—the "very large penalty" model—the judge benefits from reading the statute and rules more in line with the statute than if he hadn't read it.

However, when the legislature's penalty takes the form of a legislative override, then the judge's behavior depends on the form of the override.

For instance, under the "closest point" reversion rule, the judge doesn't benefit from reading the statute and rules perfectly in line with his ideology.

Under the "Nash bargaining" reversion rule, the judge does benefit from reading the statute and, for fear of being overridden, deviates from his ideology in the direction of compliance with the statute; however, surprisingly, reading the statute make sometimes make move *away* from the statutory point.

This shows that legislative overrides can have more complex effects on judicial behavior than has previously been understood.

Several interpretive methods When several methods of interpretation are available—and if we assume we are in the "very bad penalty" model, which does encourage statutory compliance—judges can choose which method they use, and thus they can choose which statutory point they move toward. (Because an interpretive method is a rhetorical strategy, i.e. reliance on a particular set of evidence in coming to a decision, and because a judge must "show his work," "moving toward a statutory point" means using an interpretive method in a way that is visible to the outside observer.)

Because different judges can move toward different statutory points depending on their ideology, scholars who study interpretive methods must take care not to confuse the distribution of *observed* opinions using a particular method with the *true nature* of the method, or what the universe of published opinions would look like if *every* judge were constrained to use the method.

Next steps So far, we have been assuming a naive Congress:

- The existence of two different "most plausible points" of the statute, M_1 and M_2, depending on which interpretive method is used, was assumed to be inadvertent.
- Even with a single interpretive method, we have assumed that the legislators have enacted a law at M without thinking of the judge's strategic behavior. If the legislators knew that the judge behaves as described in this paper, what would that imply as to their enactment of the law? If they knew, for instance, that the judiciary was so extremist as to push the law in a particular direction, would they enact a law that bends over backwards in the other direction? If so, can we even say that M is within the interval $[C, P]$?

We have also assumed that, when we see a point M, we know nothing except that $C < M$ and $P > M$. But if we know something about the legislators' strategic behavior (for instance, that they use Nash bargaining), perhaps we might be able to extract more information from the location of M.

The next steps in this project may include:

- Discussing how a statutory method can be more or less "constraining." Intuitively, it seems as though some interpretive methods are highly determinate while others are less so; debates over textualism, intentionalism, and purposivism often engage this question. The question in this model is what rational-choice assumptions can make methods more or less constraining.
- Discussing the implications for empirical study of judicial decisionmaking. In Volokh (2008), I argued that judges' ability to choose methods from case to case can lead to a self-selection problem, where observing existing textualist opinions can lead us to incorrect conclusions about the true nature of textualism (as it would be if, say, textualism were mandated for all judges). I anticipate that a Roy model will be the way to proceed (cf., e.g., Roy (1951), Heckman and Honoré (1990), and Borjas (1987)).

References

1. Borjas, George J. 1987. Self-Selection and the Earnings of Immigrants. *American Economic Review* 77:531–53.

2. Easterbrook, Frank H. 1988. The Role of Original Intent in Statutory Construction. *Harvard Journal of Law & Public Policy* 11:59–66.

3. Eskridge, William N., Jr. 1991. Overriding Supreme Court Statutory Interpretation Decisions. *Yale Law Journal* 101:331–455.

4. Eskridge, William N., Jr., and Philip P. Frickey. 1994. The Supreme Court, 1993 Term—Foreword: Law as Equilibrium. *Harvard Law Review* 108:26–108.

5. Eskridge, William N., Jr., Philip P. Frickey, and Elizabeth Garrett. 2007. *Cases and Materials on Legislation, Statutes and the Creation of Public Policy*. St. Paul, Minn.: Thomson/West.

6. Ferejohn, John, and Barry Weingast. 1992. Limitation of Statutes: Strategic Statutory Interpretation. *Georgetown Law Journal* 80:565–82.

7. Heckman, James J., and Bo E. Honoré. 1990. The Empirical Content of the Roy Model. *Econometrica* 58:1121–49.

8. Kennedy, Duncan. 1986. Freedom and Constraint in Adjudication: A Critical Phenomenology. *Journal of Legal Education* 36:518–62.

9. Mank, Bradford C. 1996. Is a Textualist Approach to Statutory Interpretation Pro-Environmentalist?: Why Pragmatic Agency Decisionmaking Is Better Than Judicial Literalism. *Washington & Lee Law Review* 53:1231–92.

10. Marks, Brian Andrew. 1989. A Model of Judicial Influence on Congressional Policymaking: *Grove City College v. Bell*. Unpublished Ph.D. dissertation, Washington University (St. Louis, Mo.), available at ProQuest Digital Dissertations, Pub'n No. AAT 8924401. May.

11. Marmor, Andrei. 2005. The Immorality of Textualism. *Loyola of Los Angeles Law Review* 38:2063–79.

12. Posner, Richard. 1993. What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does). *Supreme Court Economic Review* 3:1–41.

13. Rosenkranz, Nicholas Quinn. 2002. Federal Rules of Statutory Interpretation. *Harvard Law Review* 115:2085–2157.

14. Roy, A.D. 1951. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers* 3 (n.s.):135–46.

15. Volokh, Alexander. 2008. Choosing Interpretive Methods: A Positive Theory of Judges and Everyone Else. *New York University Law Review* 83:769–846.