Do We Still Need an ERA?*

Lee Epstein, Andrew D. Martin & Lisa Baldez†

May 6, 2004

*Please send all correspondence to Lee Epstein. Email: epstein@artsci.wustl.edu Post: Department of Political Science, Washington University, CB 1063, 1 Brookings Drive, St. Louis, MO 63130.

†Lee Epstein is the Edward Mallinckrodt Distinguished University Professor of Political Science and Professor of Law at Washington University; Andrew D. Martin is Assistant Professor of Political Science at Washington University; Lisa Baldez is Associate Professor of Government at Dartmouth College. We are grateful to the National Science Foundation, the Center for New Institutional Social Science, the Washington University School of Law, and the Weidenbaum Center on the Economy, Government, and Public Policy, for supporting our research. We also owe thanks to Susan Appleton for providing sage advice during the formative stages of this project; to participants in Washington University’s Workshop on Empirical Research in the Law (“WERL”), Jeffrey A. Segal, and Nancy Staudt for offering many useful suggestions; and to Scott Friedman, David Lewarchik, Shelby Johnston, Tasina Nitzschke Nikiser, and Peter Ryan for outstanding research assistance. We used R, SPSS, and Stata to conduct the analyses presented in this paper. The project’s web site [http://artsci.wustl.edu/~polisci/epstein/research/ERA.html] houses a full replication archive, including a database containing all the cases and variables we used in this study, as well as the documentation necessary to reproduce our results.
Do We Still Need an ERA?

Abstract

For over three decades, those engaged in the battle over the Equal Rights Amendment (ERA), along with many scholarly commentators, have argued that ratification of the Amendment will force U.S. courts (1) to elevate the standard of law they now use to adjudicate claims of sex discrimination, which, in turn, will lead them (2) to reach many more decisions in favor of litigants alleging discrimination. We investigate this two-step claim via an examination of constitutional sex discrimination litigation in the fifty states—over a third of which have adopted ERAs.

Employing methods especially developed and tuned for this investigation, we find, in line with the bulk of the extant commentary, that the presence of an ERA significantly increases the likelihood of a court applying a higher standard of law, which, in turn, significantly increases the likelihood of a decision favoring the equality claim. This finding, however, is not without its share of nuances. While ERAs do lead to results that their supporters desire and their opponents disdain, they are just one of several factors that explain judicial choices over standards of law and the outcomes application of those standards produce.

The practical effect of [the ERA] would be seen most clearly in court deliberations on cases of sex discrimination. For the first time, “sex” would be a suspect classification requiring the same high level of “strict scrutiny” and having to meet the same high level of justification... that the classification of race currently requires. We need the ERA to clarify the law for the lower courts, whose decisions still reflect confusion and inconsistency about how to deal with sex discrimination claims.

—Roberta W. Francis, Chair, Equal Rights Task Force of the National Council of Women’s Organizations, 2003

Without an ERA, the U.S. Supreme Court ruled that there is no right to have abortions paid for by public funds. As a result, 32 prohibit the use of state tax funds to perform abortions... But the law is different in states that have a State ERA. In New Mexico, the state supreme court... held that, since only women become pregnant or undergo abortions, the denial of taxpayer funding for them can be construed as “sex discrimination.” That argument dictates the conclusion that ERA makes taxpayer funding of abortion a constitutional right.

—Phyllis Schlafly, Founder and President, Eagle Forum, 2003

Despite pronouncements in the 1980s to the contrary, the Equal Rights Amendment (ERA) is not dead. Actually, by all indications the battle has heated up in recent years. One sign is the “three-state” strategy deployed by organized interests in response to claims appearing in scholarly papers, policy memoranda, and the press that ratification of the 27th Amendment in 1992—over 200 years after it was proposed—may hold implications, if not promise, for the ERA (see, e.g., Congressional Research Service, 1996; Duerst-Lahti, 2003; Francis, 2001; Goodman, 2000; Held et al., 1997). To be sure, this “reconstituted” drive for the ERA has generated substantial opposition (especially from Phyllis Schlafly and her Eagle Forum) but it may very well succeed in Illinois, where 65% of voters support ratification and only 19% do not (Parsons, 2003). Another sign is the increasing importance attached to the Amendment in academic and media treatments. By way of illustration consider that in the first six months of 1993, just 186 news articles made mention of the ERA; that figure for the same period, a decade later, in 2003 was well over double (N=471). It is thus hardly a surprise that ratification of the ERA has (once again) emerged as something of a hot button issue in political campaigns (see, e.g., Bousquet, 2003; Canedy, 2003; Pierce, 2003). It also is far from startling that the bulk of contemporary commentary now suggests that the ERA may be as dead as the 19th Amendment, which took over 40 years to gain ratification.

1We obtained these figures from a LEXIS search (in the news group file) on the term Equal Rights Amendment.
At a time when women “have made extraordinary gains toward establishing equality” (Lapidus, 2001), why has the ERA reemerged? Certainly some of the explanation lies in the symbolic value supporters attach to it (see, e.g., Daughtrey, 2000). But perhaps even more important are the practical consequences of the Amendment—and, indeed, it is these practical consequences that both supporters and opponents of the ERA stress. What they argue, to put their claims in general terms, is that the adoption of a formal institution—the ERA—will force U.S. courts (1) to elevate the standard of law they now use to adjudicate claims of sex discrimination, which, in turn, will lead them (2) to reach many more decisions in favor of the party alleging discrimination. Francis, in the quote above, emphasizes the first part of this belief when she writes, “For the first time, ‘sex’ would be a suspect classification.” And Schlafly underscores the second when she implies that the presence of an ERA generated an outcome that would not have resulted in its absence—largely, if not solely, because New Mexico’s equality provision prompted its tribunal to adopt a different standard of law than would (or did) the U.S. Supreme Court.

Momentarily we say more about this argument—really a two-step claim—about the possible effects of a federal ERA. For now we only wish to emphasize that altering the standard or rule used to adjudicate sex discrimination claims (that is, elevating sex to a “suspect” classification) was a primary motivation for the drive for (and against) the ERA in the 1970s (e.g., Brown et al., 1971; Dorsen and Ross, 1971; Emerson, 1971; Harvard Law Review, 1970; Kay, 1981, 1988; Kurland, 1971; Minnesota Law Review, 1973; Mezey, 1992; Voller, 1974)—and it remains so today, into the 2000s (e.g., Eskridge and Hunter, 1997; Farone, 2000; Francis, 2001; Hirczy de Mino, 1997; Kaufman, 2001; Lukey and Smagula, 2000). With an ERA, these writers maintain, the standards of law will alter sufficiently to force judges to eradicate sex-based classifications that are currently permissible, such as those that make it easier for children born out of wedlock overseas to become citizens if their mother, rather than their father, is a citizen (upheld by the U.S. Supreme Court in Nguyen v. INS 2001); those that place restrictions on abortion funding (upheld in Harris v. McCrae 1983); and those that limit the military draft to men (upheld in Rostker v. Goldberg 1981) (for other examples, see Brake, 1996; Daughtrey, 2000; Eskridge and Hunter, 1997; Freund, 1971; Gamme, 1989; Halberstam and Deseifeis, 1994; Hirczy de Mino, 1997; Hunter, 2001; Kurland, 1971).

Does this argument have merit? We raise this question because—despite the two-step claim’s resilience as a battle cry among both friends and foes of the ERA and its status as the prevailing belief among policy makers and academics alike—no one has ever systematically assessed it. Of course, determining whether a piece of conventional political wisdom can withstand rigorous scrutiny is almost always a worthwhile undertaking but it is made even more so here for, while the claim enjoys widespread support, it is not without its share of critics. Some take aim at the first part, suggesting that the ERA will not automatically and mechanically lead to the elevation of standards used to adjudicate sex-based classifications (e.g., Farone, 2000; Gammie, 1989; Kaufman, 2001); others challenge the second, asserting that even if sex becomes a suspect class, application of that higher standard will not necessarily lead to different outcomes (e.g., Case, 2000; Mansbridge, 1986; Sherwin, 1984-85).

Certainly, we understand why neither critics nor supporters of the two-step claim have attempted to verify their arguments: Because the ERA is not yet a part of the federal constitution, no one can directly observe its effect on either standards of law or case outcomes. But, as Schlafly’s words above suggest (see also Gammie, 1989; Hirczy de Mino, 1997; Linton, 1997), we can make
inferences about its impact by looking to the U.S. states. That is because over one-third have incorporated ERAs into their constitutions—with many of those amendments containing similar language and purporting to carry analogous objectives as the federal ERA (see, e.g., Crump, 1973; Farone, 2000). Hence, by examining the standards used and the outcomes produced in constitutional sex discrimination litigation in courts of last resort in states with and without sex equality provisions, we may be able to gain leverage on the potential effect of a national ERA.

At the very least, this is the task we undertake in this paper. Specifically we consider whether justices operating in states with an ERA are more likely to apply higher standards of law to adjudicate sex discrimination cases and whether application of those standards leads to outcomes more favorable to parties alleging discrimination. Our basic results, so that there will be no mystery about them, are relatively straightforward. While there may be good symbolic reasons to support (or oppose) the ERA, there are, in fact, practical grounds as well: Just as the bulk of the scholarly, policy, and interest group communities believe, the presence of an ERA significantly increases the likelihood of a court applying a higher standard of law, which, in turn, significantly increases the likelihood of a decision favoring the equality claim. These basic findings, as our emphasis on “relatively” indicates, though, are not without their share of nuances. While we do find that ERAs lead to results that their supporters desire and their opponents disdain, they are just one of several factors that explain judicial choices over standards of law and the outcomes application of those standards produce.

We develop these findings in four steps. We begin with a discussion of the conventional view of the effect of ERAs, and explain why that view may be susceptible to challenge. Next we spell out our procedures for assessing the prevailing wisdom—procedures that are attentive to the purported dependence between standards of law and outcomes, as well as to other factors that may lead judges to adopt particular standards and to reach particular conclusions. We then describe the results yielded by these procedures, along with their implications for policy debates over the ERA and, more broadly, for future disciplinary research on judicial decisions.

1 Courts, Sex Discrimination, and the ERA

The two-part argument made by many of those involved in the battle over the ERA—whether interest groups or scholars—emanates from the way courts analyze claims of discrimination under the Fourteenth Amendment’s Equal Protection Clause.\(^2\) To assess whether government actions run afoul of the clause, judges traditionally (that is, through the 1970s) applied one of two standards. Under the traditional rational basis test, as Table 1 shows, courts presume the validity of whatever classification the government has made (e.g., allowing only those over the age of 18 to enter into contracts; permitting only M.D.s to perform surgery); it is up to the party challenging the law to establish that it is irrational. Since this burden is difficult to meet, many commentators argue that rational basis leads to a predictable outcome: courts defer to the government, generally upholding its classification (see, e.g., Brown et al., 1971; Epstein and Walker, 2004; Kaufman, 2001; Mezey, 2003).

\(^2\)The Equal Protection Clause of the Fourteenth Amendment is restricted to the states; the governing constitutional provision for claims of discrimination against the federal government is the Due Process Clause of the Fifth Amendment. For purposes of our discussion on sex discrimination, the two clauses are interchangeable.
Example of Test Application Validity Standard

<table>
<thead>
<tr>
<th>Test</th>
<th>Example of Application</th>
<th>Validity Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rational Basis</td>
<td>Age discrimination</td>
<td>The law must be a reasonable measure designed to achieve a legitimate government purpose.</td>
</tr>
<tr>
<td>Intermediate scrutiny</td>
<td>Sex discrimination</td>
<td>The law must be substantially related to the achievement of an important objective.</td>
</tr>
<tr>
<td>Strict scrutiny</td>
<td>Race discrimination</td>
<td>The law must be the least restrictive means available to achieve a compelling state interest.</td>
</tr>
</tbody>
</table>

Table 1: Equal protection tests. Source: Epstein and Walker (2004, 645). See also note 4.

Until the 1970s, the vast majority of claims of discrimination proceeded under the rules of the traditional rational basis test—with one particularly relevant exception: race. In light of the history surrounding ratification of the Fourteenth Amendment, the Supreme Court has held that classifications based on race should be subject to a less surmountable standard, known as strict scrutiny (or suspect class). Under this standard, judges presume that a government action is suspect or unconstitutional; only by showing that the law is the least restrictive means available to achieve a compelling state interest can the government overcome that presumption (see Table 1). Given the difficulty of making this showing, the conventional view among scholars is that application of strict scrutiny leads to outcomes just as predictable as those under rational basis—only, of course, in the opposite direction: when courts apply this stricter test, they almost always rule in favor of the party alleging discrimination. Or, as Gunther (1971, 8) once famously put it, the suspect class test is “strict in theory and fatal in fact,” whereas the traditional rational basis standard provides “minimal scrutiny in theory and virtually none in fact.”

It is thus no wonder that as part of their effort to eradicate discrimination, women’s rights groups, beginning in the late 1960s, attempted to convince courts that sex-based classifications ought to be subject to strict scrutiny rather than to a rational basis analysis (see e.g., Daughtrey, 2000). Their litigation efforts did not succeed but neither did they wholly fail. In response to their claims, the U.S. Supreme Court in 1976 articulated a new standard—often called intermediate or heightened scrutiny—that falls somewhere in between rational basis and strict scrutiny. Under it, the challenged law must be substantially related to the achievement of an important government objective (see Table 1).

Justice O’Connor has taken issue with this claim, asserting strict scrutiny is not always “strict in theory, but fatal in fact.” This is true with regard to affirmative action programs but, for all other race-based classifications, most contemporary commentators suggest that Gunther’s assertion remains apt (see, e.g., Epstein and Walker, 2004; Farber et al., 2003; Mezey, 2003). When it comes to sex-based classifications, however, some aver that O’Connor’s view may have considerable merit. For more on this point, see Section 1.1 and note 6.

Some justices have pushed for a variation on this approach, which would require an “exceedingly persuasive justification” to sustain a sex-based classification (see, e.g., Justice Ginsburg’s opinion for the Court in United States v. Virginia [1996]). But, in light of recent Court decisions, such as Nguyen v. INS (2001), that effort has apparently...
To be sure, many argue that application of intermediate scrutiny leads to more favorable outcomes for parties alleging gender discrimination than did the traditional standard (e.g., Case, 2000; Goldstein, 1994). At the same time, though, they suggest that the intermediate approach, as opposed to rational basis or strict scrutiny, produces far less predictable results: the Court may more often than not void sex-based classifications but it more than occasionally upholds them (see, e.g., Brake 1996; Mezey 2003). Data derived from the Spaeth (2003) U.S. Supreme Court Database—showing that the party alleging sex discrimination prevailed in just slightly more than a majority of the 23 post-1976 suits (60%)—tend to support this belief, as do recent Court decisions. In United States v. Virginia (1996), the federal government invited the justices to apply strict scrutiny to sex-based classifications—an invitation that many scholars (e.g., Dorf, 2002; Sunstein, 1996), along with Justice Scalia, say the majority nearly accepted when it attempted to “ratchet up” the intermediate scrutiny standard to strike down Virginia Military Institute’s all-male admissions policy (see note 4). Just five years later, in Nguyen v. INS (2001), however, the Court upheld a federal law that privileges a mother over a father in citizenship proceedings. While the majority proclaimed that the sex-based classification created by the law achieved important government interests and, as such, passed the heightened scrutiny test, Justice O’Connor disagreed. In a vigorous dissent, she accused the Court of explaining and applying “heightened scrutiny... in a manner... that is a stranger to our precedents.” “No one,” she wrote, “should mistake the majority’s analysis for a careful application of this Court’s equal protection jurisprudence concerning sex-based classifications. Today’s decision instead represents a deviation from a line of cases in which we have vigilantly applied heightened scrutiny to such classifications to determine whether a constitutional violation has occurred. I trust that the depth and vitality of these precedents will ensure that today’s error remains an aberration.” In short, O’Connor “not so subtly implied that the majority had, in effect, not applied intermediate scrutiny, but rational basis review” (Deutsch, 2003, 187).

And therein lies the rub: Without an ERA, according to so many commentators and virtually all organized interests, the justices will continue to invoke (or perhaps ignore) the “murky” intermediate rule, upholding or voiding classifications as they see fit; and judges on state and lower federal courts will do the same or, even “concoct” their own approaches to sex discrimination (see, e.g., Brake, 1996; Francis, 2001; Mezey, 2003). But with an ERA jurists will be forced (1) to elevate sex to a suspect class, which in turn will lead them (2) to eradicate virtually all sex-based classifications (i.e., not just 60%), as they now do in the case of race.

1.1 Plausibility of the Two-Step Claim

This last sentence encapsulates what we have labeled the two-step claim about the ERA’s impact—and it is certainly not one lacking in support or plausibility. Mounds of scholarly commentary point to its merit (see, e.g., Brown et al., 1971; Dorsen and Ross, 1971; Emerson, 1971; Eskridge and Hunter, 1997; Farone, 2000; Ginsburg, 1978, 1979; Harvard Law Review, 1970; Kaufman, 2001; Kay, 1981, 1988; Lukey and Smagula, 2000; Minnesota Law Review, 1973; Vollers, 1974); friends and foes of the ERA alike post it on their web sites (see, e.g., Eagle Forum 2003; National Council of Women’s Organizations 2003); the history of race discrimination litigation portends it (see, e.g., Farber et al., 2003; Greenberg, 1976); and even U.S. Supreme Court justices have suggested as much; indeed, in the early 1970s several declined to elevate sex to a suspect class at least in part because they thought it “inappropriate to ‘amend’ the Constitution while the ERA was pending” failed (see, e.g., Bowsher, 1998; Stobaugh, 2002; Weinrib, 2003).
(Eskridge and Hunter, 1997, 78; see also Ginsburg 1978, 1979).

Even so, we hasten to note, the two-step claim is not without its share of detractors. Some take aim at the first part, explicitly asserting that the presence of an ERA does not ensure adoption of strict scrutiny and that its absence does not automatically negate it (Brake, 1996; Denning and Vile, 2000; Gammie, 1989; Kaufman, 2001; O’Neill, 1993; Sherwin, 1984-85). In support of this belief commentators point to federal courts that have all but ignored the current intermediate standard and have instead invoked higher or lower rules as they so desire; they also point to state courts of last resort that have used the intermediate or even rational basis standard to adjudicate sex discrimination cases even if their constitution contains an ERA, as well as to supreme courts that invoke strict scrutiny in the absence of an ERA. Many more commentators, though, challenge the second part of the claim, averring that strict scrutiny will not always be “fatal” to sex-based classifications. Leading the way here is Mansbridge (1986) who argues that the ERA “would almost certainly have made sex at least a ‘suspect’ classification,” but that “treating sex merely as a ‘suspect’ classification would [leave judges] free to strike down any statute involving sex discrimination that did not in their view have a compelling justification, while leaving them free to uphold any statute that they thought did have compelling justification” (see also Case, 2000; Mezey, 2003; Schoen, 1978; Sherwin, 1984-85; Stanford Law Review, 1950; Tarr and Porter, 1982).

What these analyses underscore, to frame their critiques more broadly, is an argument advanced in study after study of judicial decisions; namely, while institutions—whether formal (such as constitutional dictates) or informal (such as legal principles)—may “matter,” they are not as determinative of outcomes as the two-step claim makes them out to be. Indeed, the extant literature on judging typically defines institutions as sets of rules that structure interactions (see, e.g., Epstein and Knight, 1998; Murphy, 1964), not as rules that establish outcomes; and it typically views the choices judges make as a function of many other forces, including the judges’ own political preferences and personal attributes, features of the external environment in which they deliberate, and the characteristics of the particular suits they must resolve (see, e.g., Caldeira et al., 1999; Cross and Tiller, 1998; George and Epstein, 1992; Gryski et al., 1986; Pinello, 2003; Revesz, 1997; Segal et al., 1995).

In emphasizing this point, we do not mean to suggest that scholars of judicial politics believe formal and informal rules to be irrelevant. While a number may subscribe to this view (or, at the very least, do not build institutions into their decision-making models), many others do not—as exemplified by studies arguing that particular types of institutions for selecting judges may be more likely than others to induce sophisticated judicial behavior (e.g., Bright and Keenan, 1995; Burbank and Friedman, 2002; Tabarrok and Helland, 1999) and those explaining why formal constitutional dictates and the norm favoring respect for precedent may serve to constrain judges, even those who prefer different rules (e.g., Caminker, 1994; Knight and Epstein, 1996). What we do mean to suggest rather is that in none of these studies, or in any others we can identify, are

5 Of course, federal courts are supposed to adhere to legal principles established by the U.S. Supreme Court; and state courts are supposed to view federal law as establishing a floor (though not a ceiling) on civil rights and liberties below which they cannot go. But, as the numerous studies we cite in the text—not to mention our own reading of the cases—indicate, these norms do not always hold, at least not in this area of the law. Even the U.S. Supreme Court occasionally departs from its own standard to adjudicate sex discrimination cases (see e.g., Mezey, 2003).

6 These (largely doctrinal) analysis indicate that (1) courts applying the same standard to assess the same sex-based classification at roughly the same point in time do not always reach the same conclusion and (2) courts applying rational basis do not always uphold sex-based classifications and courts applying suspect scrutiny do not always strike them.
the authors contending (in contrast to the two-step claim) that institutions determine outcomes in all circumstances and in all matters. In fact, a central message in the extant literature on judging (see especially George and Epstein, 1992; Pinello, 2003; Segal et al., 1995) and the basic point of Mansbridge’s analysis of the potential effect of an ERA is that rules may constrain courts, but they do not—as so much commentary on the ERA seems to maintain—mechanically lead them to make particular choices; many other factors come into play.

2 Data and Research Methods

These sorts of critiques, not to mention the lack of empirical scrutiny, underscore the importance of assessing the two-step claim about the impact of the ERA. While there may be good reasons to believe that ERAs will lead to the adoption of higher standards, which in turn will generate outcomes more favorable to parties alleging discrimination, there are equally good reasons to question that belief. At the very least, the two-step claim, so deeply embedded in the literature on sex-based classifications, strikes us and others as asking too much of formal and informal rules. While rules certainly can serve to structure choices, it seems imprudent to believe that they do all the work—especially when so many studies of judging suggest otherwise.

Of course, we can assess these various perspectives empirically. As we noted at the onset, an investigation into the standards used and outcomes produced in constitutional sex discrimination cases in nation’s 50 state courts of last resort—that is, those operating in legal environments with and without an ERA—should permit us to draw inferences about the validity of the two-step claim. So too such an investigation ought to facilitate assessments of the argument of some, that an ERA will not necessarily lead to the application of strict scrutiny; and of others, that application of strict scrutiny will not necessarily lead to predictable, pro-equality outcomes; that rather in both instances other factors play a crucial role explaining judicial choices. And while our assessment of these claims, as well as any inferences we may draw about the effect of an ERA, will be of a higher (i.e., more certain) quality for the context we can observe—the states—than for the one we cannot observe—the federal level, in light of the methodological tack we take, which entails controlling for many factors relevant to state and federal judging, not to mention the similar language and purposes of the state and national amendments, we should be able to gain some leverage on the potential effect of a national ERA.

It is to this methodological approach that we now turn. Specifically, in what directly follows, we describe the procedures we used to gather data on our unit of analysis (constitutional sex discrimination cases), the variables we incorporated into our statistical model, and the method we employed to estimate the model. Section 3 details the results yielded by these procedures.

2.1 The Unit of Analysis: Constitutional Sex Discrimination Cases

To assess the effect of an ERA, we set out to collect data on all constitutional sex discrimination cases resolved in state courts of last resort between 1960 and 1999. This proved a challenging task.

---

7For the sake of convenience, throughout this paper we use the terms “state court of last resort” and “state supreme court” interchangeably even though we recognize that some state courts of last resort are not named “supreme court.”

8We settled on this time frame for reasons practical (e.g., the existence, as we explain later in the text, of valid and reliable measures of judicial preferences), analytical (e.g., a sufficient time line pre- and post-ratification of most state ERAs) and jurisprudential (e.g., an ample period since the instantiation of contemporary equal protection doctrine) in nature.
No equivalent of Spaeth’s (2003) U.S. Supreme Court Database exists for state supreme courts; and searches of electronic resources (such as LEXIS or Westlaw) are unlikely to produce all the relevant suits.

In light of these obstacles, scholars investigating state court decisions have developed their case lists from multiple sources. In amassing his database on gay rights litigation, for example, Pinello (2003) conducted computer-aided searches and drew on scholarly writings on the subject. We generally followed suit. In our quest to uncover all constitutional sex discrimination cases, we too performed various LEXIS searches and systematically extracted cases from extant literature on the subject.⁹ Heeding the lessons of research by Gryski and his colleagues (1986), however, we supplemented these sources by “shepardizing” the U.S. Supreme Court’s major constitutional sex discrimination cases.¹⁰

After perusing each opinion unearthed by these various procedures—mainly to ensure that the court in question was the state court of last resort, that the dispute was real and not hypothetical, and that, in the course of resolving the suit, the state justices addressed a claim of constitutional sex discrimination¹¹—we were left with 416 cases.¹² Of course, we cannot say that we identified each and every pertinent case but, like Pinello (2003), we do not believe we missed many. We also can say that our procedures are reproducible (all the data and documentation are available on our web site), replicable, and capable of updating and backdating.

2.2 Variables Incorporated into the Analysis

With the 416 cases in hand, we set out to collect the information necessary to animate the three primary variables of interest: (1) the presence or absence of an ERA; (2) the standards courts use to adjudicate sex discrimination claims, and (3) case outcomes. We also amassed data designed to assess the various factors suggested by previous studies of judging that may affect the standard used, the case outcome, or both. We elaborate on these below.

2.2.1 Variables of Primary Interest: An ERA, the Standard of Law, and the Outcome

The lynch pin of the two-step claim is the presence (or absence) of an ERA. It is (1) an ERA, under the claim, that necessitates the use of a higher standard of law, and it is (2) the application of a higher standard of law that generates an outcome favorable to the party alleging discrimination. To put it another way, in the first part of the claim the ERA is the key causal variable and the standard

---

⁹The project’s web site [URL omitted to protect the authors’ identity] provides the details of these searches, including the terms we entered into LEXIS, to identify the cases and the literature. Note to referees: To facilitate your review of this manuscript, we provide information about our searches in Appendix C, note 9.

¹⁰Shepard’s (now available via LEXIS/NEXIS Academic Universe) enables researchers to identify all cases citing to a particular case. A list of the 13 Court cases we shepardized is available our web site. Note to referees: To facilitate your review of this manuscript, we identify the cases in Appendix C, note 10.

¹¹Specifically, we excluded courts in the District of Columbia and Puerto Rico and eliminated the handful of advisory opinions produced by our searches because not all state courts have the authority to issue these sorts of opinions. In light of the purpose of our project, an additional criterion was that the court invoked a standard of law (see Table 1) to adjudicate the sex-based claim. Since we return to this point later (see Section 2.2.1), suffice it to note here that in secondary analyses designed to assess the robustness of our results we incorporated those cases in which the court failed to apply a standard (by adding the variable 0 to the standard variable; see Section 2.2.1). The results, which are housed on our web site, were remarkably stable and consistent with those depicted in Table 2 (see Section 3).

¹²The project’s web site houses a full replication archive, including a database containing all the cases and variables we used in this study, as well as the documentation necessary to reproduce our results.
is the dependent variable; in the second part, the standard serves as the key causal variable, and the outcome, the dependent variable.

Measuring the first independent variable, the presence of an ERA, is a relatively straightforward task, requiring us to do little more than determine whether an ERA was in effect (coded as 1) or not (coded as 0) at the time the court resolved the case.\textsuperscript{14}

Turning next to the standard of law, which serves as both a dependent and an independent variable in our study, we code each into one of the three categories depicted in Table 1: 1=ratio-nal basis (minimal scrutiny), 2=intermediate (heightened) scrutiny, or 3=strict scrutiny (suspect class).\textsuperscript{15} Since our replication archive houses a detailed coding instrument, suffice it to mention here the primary problem we encountered: cases in which we could not identify a standard because either the court simply did not apply or articulate one or because it explicitly refused to select among various alternatives. We exclude these cases from our primary analysis (that is, they are not among the 416) but we include them (N=51) in supplemental investigations designed to assess the robustness of our results.

Finally, we required information about the outcome of the litigation. Here we take the approach commended by Gryski et al. (1986), among others, and code whether the party alleging sex discrimination won (= 1) or lost (= 0) the dispute.\textsuperscript{16}

\subsection*{2.2.2 Other Factors Influencing Judicial Decisions}

These three variables—(1) the presence or absence of an ERA; (2) the standards courts apply to sex discrimination claims, and (3) case outcomes—provide us with the information essential to scrutinize the two-step claim. What we lack are the data necessary to assess challenges to that specific claim or, for that matter, any claim that emphasizes the utter determinacy of rules. Examining such challenges requires us to attend to the other forces that come to bear on judging—specifically those that may influence the two dependent variables of our study, the standard of law and the case outcome. From a close reading of the literature, as we allude above, emerge two sets of factors relevant to the choice of standard: features of the state supreme court and of the...

\textsuperscript{13}A list of the 18 “ERA states” along with the date and exact wording of their sex equality provisions is available on our web site. \textit{Note to referees:} See Appendix C, note 14.

\textsuperscript{14}As our emphasis on “relatively” indicates, we encountered several (fairly minor) problems in coding this variable. We describe these on the project’s web site, as well as the steps we took to account for them in our analyses. \textit{Note to referees:} To facilitate your review of this manuscript, we provide information about these problems in Appendix C, note 14.

\textsuperscript{15}The project’s web site houses our coding rules, along with the results of our reliability analysis on the coding of this variable. \textit{Note to referees:} To facilitate your review of this manuscript, we provide this information in Appendix C, note 15.

\textsuperscript{16}We are well aware of normative debates among some feminists over whether, as Goldstein (1994, 209) puts it, “to argue for protective legislation for women on the grounds that without such legislation women are unfairly disadvantaged by making them play by rules that were designed with men in mind, and that are ill-adapted to women’s biology and life patterns.” While we appreciate this argument, our coding scheme remains relatively agnostic over it (\textit{e.g.}, if a pregnant woman fired from her job is the plaintiff in a suit alleging sex-based discrimination in violation of her constitutional rights, she would be the party alleging sex discrimination.) We also understand the position of some scholars that a formal rule, such as an ERA, will not effectively end the subordination of women by men at least in part because of the “problem of male dominance in virtually all facets of social, political, and economic life” (MacKinnon 1987; see also Becker, 1989; Case, 2002). To be sure, we do not attempt to assess this position but our analysis does lend support to the claims of Sullivan (2002) and others who argue that formal equality provisions are not always inefficacious but rather their effectiveness depends a good deal on whom is interpreting them. Specifically, to foreshadow our results, we find that as the fraction of women serving on a state supreme court increases, the likelihood of the court adopting a higher standard of law also increases—and significantly increases at that.
external socio-legal environment in which it operates. Some of the same factors also may shape the outcomes of the disputes but, to account fully for them, we also must consider an additional set: specific characteristics of sex discrimination suits.

Beginning with features of the court, existing research emphasizes two. One is the political preference of the majority of its members, as structured by the institutions used to select them. This last clause is crucial for, again as we noted earlier, many analysts argue that particular types of selection systems are more likely than others to induce sophisticated behavior on the part of actors—such that the greater the accountability established in the institution, the higher the opportunity costs for judges to act sincerely, and, accordingly, the more extensive sophisticated behavior will be (e.g., Bright and Keenan, 1995; Burbank and Friedman, 2002). On this account, then, the decisions of elected judges will be more in line with popular sentiment, and those of appointed judges, more on track with the preferences of the governmental regime (see, e.g., Croly, 1995; Dahl, 1957; Gryski et al., 1986; Pinello, 1995; Tabarrok and Helland, 1999). Standard partisan-based measures of the judicial preferences—such as the proportion of Democrats on the bench—are thus inappropriate for this reason, and for another unique to our study: During the years under analysis here, the political parties switched positions on the ERA (Wolbrecht, 2000), thereby making any proxy of preferences that relies primarily on partisanship (including party-adjusted ideology scores) even more dubious for our purposes. But there is one measure that nicely fits our needs: the government and citizen ideology scores Berry et al. (1998) developed and assigned to each state for each year between 1960 and 1999.17 Operating under an assumption common in many studies—that we can assess the preferences of state justices on the basis of the ideology of the citizenry (if they are elected via partisan or non-partisan ballots) and on the basis of the government (if they are appointed)—we assigned a Berry et al. score to each case in the database, depending on the selection system in effect at the time the court resolved the dispute. Given the way Berry and his colleagues computed the scores (with higher numbers representing higher degrees of liberalism), we expect a positive relationship between the scores and standards: the higher the score (that is, the more liberal the court), the more likely it is to apply a higher standard of review; we also anticipate that liberal courts will reach more equality-oriented outcomes.

A second feature of the court worthy of inclusion centers on the presence of women justices. Virtually from the day Sherry (1986) penned her classic work on the possibility of a “feminine” jurisprudence scholars have hotly debated whether female judges “speak in a different voice” (for recent reviews of this literature, see, e.g., Epstein and Mather, 2003; Kay and Sparrow, 2001; Schneider, 2001). While the results of various research projects exploring judicial votes are decidedly mixed, those centering on jurisprudence (Sherry’s original target)—especially in the area of sex discrimination—are clearer. A consensus now exists that women have “pushed the law forward in sex discrimination cases” (Kay and Sparrow, 2001), with their distinct approach to legal principles possibly altering the choices made by their of male colleagues (Sherry 1986; Sullivan 2002). Based on this consensus,18 we think it reasonable to hypothesize that the greater the fraction of female justices

---

17Berry et al. (1998) reports scores for the period between 1960 and 1993. We used the updated and revised scores (through 1999) (“revised 1960-1999 government/citizen ideology series”) that Berry and his colleagues deposited with the ICPSR (Study No. 128). By necessity, partisanship is a factor—albeit one among many—that figures into the government measure.

18Note that because this consensus is over jurisprudence and not votes, our hypothesis pertains to standards, and not outcomes. We did, however, assess whether the fraction of female justices impacts outcomes. As it turns out, the variable produces a positive but insignificant coefficient—though in no way affects the basic findings reported in Table 2. Referees: See also Appendix C, note 21.
justices on the court, the greater the probability of the adoption of a higher standard of law.\textsuperscript{19}

These variables relate to features of the court; a second set pertains to the socio-legal environment in which the court operates. A relevant factor along these lines is whether an intermediate appellate court existed at the time the court decided the case (coded 1 if it existed; coded 0 otherwise). Scholars include this variable in almost all models of state court decisions out of the belief that appellate tribunals “siphon off routine cases,” leaving the more serious matters for courts of last resort to resolve. In the context of our study, the presence of these courts thus should increase the likelihood of the court applying a higher standard of review, as well as finding in favor of the party alleging discrimination, “since frivolous appeals will be screened by the intermediate appellate courts, and the high courts will be evaluating only the more serious claims of sex discrimination” (Gryski, Main, and Dixon 1986, 145; see also Atkins and Glick 1976; Canon and Jaros 1970).

A second variable centering on the socio-legal environment is the state’s level of enthusiasm for the national ERA, which we also can capture with a simple dichotomous variable: whether the state ratified the amendment (=1) or not (=0). Because this may provide important information to the justices about the public’s receptivity to the application of a higher standard of law to resolve sex discrimination cases, it may very well exert an influence on that choice quite apart from the presence (or more pointedly, absence) of a state ERA. At the very least, ratification of the federal ERA could send state justices precisely the opposite signal that the failure to ratify sent to their federal counterparts; namely, they would not be usurping their state’s political process if they elevated sex to a suspect class (Eskridge and Hunter, 1997; Ginsburg, 1978). Accordingly we expect justices operating in states that ratified the national ERA to apply a higher standard of review.

As our discussion so far suggests, we foresee that various features of the courts and of the environment under which they operate will affect both the standard they apply and the outcome they reach. There is yet a third set that we anticipate will influence only the latter: characteristics of the litigation. The general idea here is that even if the justices adopt a particular standard of law to adjudicate sex discrimination claims, they will apply it to reach different outcomes depending on the features of the case (see, e.g., Beck and Baker, 1993; Mansbridge, 1986; McCausland, 1983; Mezey, 2003; Simpson, 1977).

Operating under this logic, we combed the extant normative and doctrinal literature to identify relevant characteristics of the litigation that may affect case outcomes. Three emerged in virtually every analysis: whether or not (1) the government defended a sex-based classification, (2) the government or another party defended the classification on the basis of a relevant physical difference between men and women, and (3) a woman claimed discrimination. Our expectation about the government’s involvement is simple enough: when, as a party or an amicus curiae, it defends a sex-based classification, we expect the likelihood of defeat for the litigant alleging a violation of his or her rights to increase. This hypothesis reflects a voluminous literature both in the federal and state judicial contexts demonstrating that courts defer to the government—especially in criminal cases (see, e.g., Atkins and Glick, 1976; Emmert, 1992; Glick and Vines, 1973; Gryski et al., 1986; Traut and Emmert, 1998; Wheeler et al., 1987; Yale Law Journal, 1978). For our purposes, we simply incorporate a variable indicating government involvement (coded 1) or not (coded 0) but we also, in a supplementary analysis, control for the type of suit (criminal or civil).

Turning to the second case characteristic, scholars are virtually unanimous in their belief that

\textsuperscript{19}Allen and Wall (1993) contains data on women serving on state courts of last resort between 1922 and 1992; we updated this information from various sources, including court web sites.
a defense of discrimination on the basis of a relevant physical difference makes it more difficult for both federal and states courts to resolve the dispute in favor of the party alleging discrimination—regardless of the standard of law they use (see, e.g., Epstein and Walker, 2004; Kay, 1981; Mezey, 2003). Hirczy de Mino’s (1997) doctrinal analysis of the Texas high court decisions, for example, finds that even when the justices apply an elevated standard they tend to rule against the equality claim in litigation involving a physical difference (he specifically points to toplessness). Mezey (2003) and many others make a similar point about federal jurists, explaining that some have downright asserted that courts ought apply a different standard when such differences arise. The resulting hypothesis, then, is straightforward: if a defense based on a relevant physical distinction arises in the litigation (coded 1 if yes; coded 0 if no), we anticipate that the justices will find against the party alleging discrimination.

Far less straightforward is the role of women litigants in increasing or decreasing the probability of success for the party alleging inequity. In perhaps the most systematic analysis of litigation in this area, Gryski and his colleagues (1986) found that state tribunals are more sympathetic to claims of discrimination made by females as opposed to males, and Mezey (2003) has suggested much the same of the U.S. Supreme Court. Given the perceived purpose of an ERA, these results might not seem particularly surprising but some observers have balked. Various feminist legal scholars, for example, have argued that the male-dominated judiciary manipulates seemingly equally-oriented principles to reach outcomes that further the interests of its “own kind” (see note 16); and doctrinal analyses of cases in Texas and Pennsylvania, two states with an ERA, suggest the plausibility of this view (Hirczy de Mino, 1997; Beck, 1993). These conflicting findings make hypothesizing difficult; yet, in light of the more systematic evidence marshaled in the Gryski et al. and Mezey studies, we expect the probability of success for the party alleging discrimination to increase when that party is a female (coded 1; coded 0 if not a female).

2.3 The Statistical Model and Estimation Method

As we have just outlined it, our analysis incorporates four variables, in addition to the presence of an ERA, to account for the standard of law applied by the court; and five, in addition to the standard of law, to account for the outcome. Since Appendix A provides a summary of these variables, we need only mention here that, given our coding scheme, we expect all variables (the presence of an ERA, the preferences of the court, the proportion of the court composed of women, the existence of a state intermediate appellate court, and national ERA ratification) to be positively related to the standard of law; we expect four (the standard of law, the preferences of the court, the existence of an intermediate appellate court, and a female as the equality party) to be positively related to case outcome and two to be negatively related (a claim of a physical difference between men and women and the government as a defender of the sex-based classification).20

Assessing these expectations, whether flowing from the two-step claim about the ERA’s impact or from criticisms of it, presents a technical challenge of no small proportions: Because we aim to explore the assumption that the choice of standard and the outcome are dependent, we must analyze two equations, with two differentially measured dependent variables—standard of law (an ordinal variable) and outcome (a dichotomous variable)—in one model. Since no standard statistical model will adequately perform this task, we developed one: a bivariate mixed response probit model, which

---

20 These represent the variables in (and concomitant expectations of) our primary analyses. As we have noted throughout, we consider variations in an effort to assess the robustness of our results. See Section 3 for the details.
allows for correlation across two equations and which we estimate using maximum likelihood.

Appendix B provides the details of this statistical model, along with our estimation methods. What is important here is that even though the parameter estimates resulting from these procedures admit to an interpretation akin to probit coefficients, our methodological approach is distinctive in two regards. First, it enables us to estimate parameters that—while substantively similar to those that would result from analyzing decisions over standards of law and case outcomes independently—are more efficient because we employ all the data to obtain them. Second, the approach facilitates a more exacting investigation of the dependence between the choices of standard and outcome (i.e., the dependence assumed by the two-step claim) because we are able to obtain a precise estimate of that dependence (in the form of an estimate of a correlation parameter, \( \rho \)) as a result of our ability to control for the factors that may affect both the standard and outcome in one model.

3 Results of the Analyses

Estimating the bivariate mixed response probit model leads to the results depicted in Table 2—results that are quite striking: All the variables produce statistically significant coefficients, and in the expected direction. The \( \rho \) estimate, indicating the correlation between the standard of law used and the outcome reached, also attains statistical significance.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 ) Constant</td>
<td>-1.106</td>
<td>0.189 *</td>
</tr>
<tr>
<td>( \beta_1 ) ERA</td>
<td>0.460</td>
<td>0.120 *</td>
</tr>
<tr>
<td>( \beta_1 ) Judicial Ideology</td>
<td>0.011</td>
<td>0.004 *</td>
</tr>
<tr>
<td>( \beta_1 ) ERA Ratification</td>
<td>0.456</td>
<td>0.143 *</td>
</tr>
<tr>
<td>( \beta_1 ) Proportion Women</td>
<td>2.187</td>
<td>0.566 *</td>
</tr>
<tr>
<td>( \beta_1 ) Intermediate Appellate Court</td>
<td>0.354</td>
<td>0.138 *</td>
</tr>
<tr>
<td>( \tau_1 ) Cut Point</td>
<td>1.179</td>
<td>0.084 *</td>
</tr>
<tr>
<td>( \beta_2 ) Constant</td>
<td>-0.794</td>
<td>0.225 *</td>
</tr>
<tr>
<td>( \beta_2 ) Judicial Ideology</td>
<td>0.017</td>
<td>0.004 *</td>
</tr>
<tr>
<td>( \beta_2 ) Intermediate Appellate Court</td>
<td>0.376</td>
<td>0.153 *</td>
</tr>
<tr>
<td>( \beta_2 ) Physical Difference</td>
<td>-0.685</td>
<td>0.160 *</td>
</tr>
<tr>
<td>( \beta_2 ) Government Opposition</td>
<td>-0.725</td>
<td>0.130 *</td>
</tr>
<tr>
<td>( \beta_2 ) Female</td>
<td>0.376</td>
<td>0.138 *</td>
</tr>
<tr>
<td>( \rho ) Correlation</td>
<td>0.532</td>
<td>0.060 *</td>
</tr>
</tbody>
</table>

Table 2: Maximum likelihood estimates and (asymptotic) standard errors for the bivariate mixed response probit model fit to the state constitutional sex discrimination data. \( N = 416 \). \( \ln L = -593.0884 \). * denotes statistical significance (\( \alpha = 0.05 \)).

Taken collectively, these basic results lend support to the two-step claim about the ERA’s impact—but not without important caveats. In what follows we explore those caveats, which emerge from an examination of the substantive impact of our findings. But two points deserve emphasis
before we proceed further. The first is that while Table 2 reflects our primary thinking about the impact of state ERAs, a sufficient number of concerns about our measures and specification—many of which we raised in Section 2—led us to undertake a full-fledged assessment of the robustness of our principal results.\textsuperscript{21} Our web site contains a reckoning of this assessment (see also note 21) but the main point here is that it leads to no changes in the interpretation we offer below: in none of these alternative specifications did the parameter estimates for the ERA and $\rho$ fail to reach statistical significance.

A second point also relates to the integrity of the results displayed in Table 2 but centers on a potential problem caused by our data (rather than our measures)—specifically that they cluster by court and time. While an insufficient amount of data exists to model extra court or time dependence, we went to some length to ensure that our inferences were not unduly affected by possible clustering. Specifically, we recomputed our standard errors (clustering by court, decision year, and court within decision year) using the White (1980) sandwich estimator, which generates standard errors that remain consistent even in the face of unmodeled dependence due to clustering.\textsuperscript{22} When we invoke these standard errors, rather than the asymptotic ones reported above, the results do not change markedly;\textsuperscript{23} most important, the key variables of interest (an ERA and the standard of law) are not affected whatsoever. Accordingly, for the substantive interpretation to follow in Section 3.1, we rely on the regular (asymptotic) standard errors.

3.1 The Two-Step Claim

Our diagnostic work noted, let us turn to a deeper consideration of the results, beginning with the prevailing wisdom encapsulated in the two-step claim about ERA’s effect. As to the first part of the claim, the findings generally are supportive, but not without their share of nuances. On the one hand, as we can observe in Table 2 the coefficient produced by the ERA variable is positive and significant, indicating that the presence of such a formal equality provision does in fact increase the probability of the adoption of a higher standard of law. So too, as we can see in Figure 1, that statistical finding is not without substantive import. What the figure shows is that the probability, over the thirty-year period, of the application of a higher standard of law varies considerably depending on whether the state adopted an ERA: When we set all other variables at their mean, the likelihood, on average, of a court invoking strict scrutiny to adjudicate a sex-based claim is just .11 in the absence of an ERA. That probability doubles in the presence of an ERA, to .23. On the other hand, because the figure of .23 is relatively distant from 1.00, it is far from certain that an ERA will lead to the automatic application of strict scrutiny. The presence of an ERA may raise the probability of courts taking that step, but it does not—just as some scholars have contended and as more general work on judicial politics would suggest—assure it.

\textsuperscript{21}Note to referees: To facilitate your review of this manuscript, we provide information about this assessment in Appendix C, note 21.

\textsuperscript{22}See Appendix B for details; the robust standard errors are included in the replication archive. Note to referees: See also Appendix C, note 22.

\textsuperscript{23}Specifically, when clustering by year, all coefficients remain statistically significant; when clustering by court, the Intermediate Appellate Court variable becomes insignificant in the standard equation; when clustering by court within year, the Government Opposition and Female variables become insignificant in the outcome equation.
Figure 1: Kernel density estimates of the probabilities of strict scrutiny, intermediate scrutiny, and rational basis as affected by the presence or absence of a state Equal Rights Amendment, accounting for parameter uncertainty. These were computed using a CLARIFY-like simulation (see King, Tomz, and Wittenberg 2000). The dashed line represents states without an ERA; the solid line represents those with an ERA. All other variables are held fixed at their sample means.

If the results pertaining to the first part of the two-step claim are somewhat mixed, those on the second part are far cleaner and stronger. The $\rho$ coefficient in Table 2 indicates that the standard a court uses and the outcome it reaches are significantly correlated; and the probabilities displayed in Figure 2 reveal that the relationship is substantively meaningful as well. Notice the monotonic increase in those odds, such that when courts assess sex classifications via a rational basis test—the lowest level of scrutiny—the likelihood of finding in favor of the equality claim is just .20. That probability increases to .47 when courts apply intermediate scrutiny and to .73 when they invoke strict scrutiny. In other words, and in line with the bulk of the extant literature, under mid-level scrutiny litigants alleging sex discrimination are nearly as likely to win as they are to lose, while application of the lowest and highest standards leads to rather predictable outcomes—though in opposing directions: under a rational basis standard, claims of sex discrimination will, on average, fail and under strict scrutiny they will, in all likelihood, prevail.
3.2 Features of the Court, the Environment, and the Cases

Taken collectively, it would hardly be a stretch to read our results as providing further ammunition to all those engaged in the debate over the ERA—friends and foes alike—as well as to many of the scholars who have commented on the Amendment’s likely effect. But neither they nor we should ignore the full lessons of our modeling exercise; namely, the presence (or absence) of a formal constitutional provision is not the only variable exerting an influence on the choice of standard; and the use of a high (or low) standard of law is not alone in accounting for case outcomes.

Turning first to the decision over what standard of law to apply, Table 2 shows that all the incorporated variables play some role in explaining that choice. The fraction of women on the bench holds particularly impressive explanatory power. As that fraction increases, as Figure 3 illustrates, the probability of applying a higher standard of law soars, even after controlling for the presence of an ERA. To see the magnitude of the effect, consider a court composed exclusively of male justices. On average, the odds of that court using a rational basis standard, setting all other variables at their mean, is a hefty .50; the probability of that same court applying strict scrutiny is but .12. Now consider a court nearly equally divided between male and female judges: as the figure illustrates, the probabilities nearly reverse: the odds of this court applying rational basis are (on average) but .14, while the probability for strict scrutiny jumps to .47.

Figure 2: Kernel density estimates of probabilities of an outcome favoring the litigant alleging sex discrimination given the rational basis standard (the left-most dashed line), the intermediate standard (the middle dashed line), and the strict scrutiny standard (the solid line). These estimates account for all parameter uncertainty, and were constructed from the simulation outlined in Appendix B. All covariates are held at their sample means.
This finding lends empirical support to the writings of Sherry (1986) and others on the existence of a feminine jurisprudence. While women may not speak in a different voice in all legal areas, nor consistently vote differently from men, their presence on the bench seems to exert an influence on how their colleagues structure the adjudication of sex-based claims. Similarly interesting is the importance of the sex of the equality-oriented litigant in explaining case outcomes. As we can observe in Table 2, this variable attains statistical significance; and it too has a meaningful influence on who ultimately prevails in the suit. If we set all other variables at their sample mean, the probability of the court finding discrimination is nearly .50 when a woman brings the suit; it dips to about a third for all other litigants. In light of the history of ERAs, not to mention the findings of empirically based studies exploring the outcomes of sex discrimination litigation (see, e.g., Gryski et al., 1986; Mezey, 2003), this finding is hardly a surprise; yet it does present something of a challenge to (largely) normative scholarship arguing that men have been the largest beneficiaries of equality provisions.

Finally, notice the role played by the political preferences of the judges—both in terms of
accounting for the standard of law applied and the outcome reached. The variable reaches statistical significance in both equations (see Table 2): As the judges’ preferences move from right- to left-of-center, the odds of applying a higher standard and of reaching a decision in favor of the litigant claiming discrimination increase. Figure 4 substantiates this point with regard to case outcomes: the probability of an outcome favoring the equality claim jumps from 0.15 to 0.74, as courts move from extreme conservatism to extreme liberalism and when we hold all other variables at their means.

Indeed, the effect of ideology (along with the proportion of females on the court, the sex of the litigant, and so on) is so dramatic that we might question the relative substantive importance of the two key variables of interest here—the presence or absence of an ERA and the standard of law. How likely is it, in other words, that courts will apply the highest legal standard to adjudicate sex discrimination claims in the absence of an ERA but in the presence of other relevant explanatory factors?

As it turns out, the effect of an ERA is hard, though not impossible, to minimize. To see this, consider a court that is extremely liberal (100 on the Berry et al. measure) and that has the highest proportion of female judges in our sample (about .50). If that court operates in a state without an ERA, the probability of it adopting strict scrutiny is a fairly high .64; in the presence of an ERA, the likelihood increases, though certainly not precipitously, to .79. Interesting too is that the .79 probability drops rather dramatically—to well below .64—if either the proportion of females or the court’s liberalism dips to or below mean levels. So, for example, even if a court in an ERA state is composed of about half females, the probability of it adopting strict scrutiny drops to .59 (from .79) should it be composed of ideological moderates (i.e., a court at the ideological mean). The
comparable figure for the same court in a non-ERA state is .40.

What these various scenarios show is that, under the “right” circumstances (say, an extremely liberal court with an unusually high proportion of women), the lack of an ERA will not necessarily stand as a barrier to the adoption of strict scrutiny. That is why we say it is not impossible to minimize an ERA’s effect. But it is also why we say it is difficult to do so: the “right” circumstances are far from the normal circumstances (they are instead the extreme and unusual circumstances). And, even when they do occur, they never generate probabilities of the adoption of strict scrutiny that are higher than those obtained in states with ERAs. Under every possible scenario (e.g., setting both ideology and the proportion of women at their means; setting ideology at the lowest level and women at the highest, and so on) courts in ERA states are more likely to adopt strict scrutiny than those operating in states lacking the Amendment.24

4 Discussion: Do We Still Need an ERA?

In the title of this paper, we ask a straightforward question: Do we still need an ERA? With our analyses now complete, we can supply an equally straightforward answer: If we believe that it is desirable for courts to produce a larger number of equality-oriented outcomes, then an ERA will not impede that objective. Quite the opposite: The presence of an ERA, even after controlling for other relevant factors, increases the probability of a court applying a higher standard of law to adjudicate claims of sex discrimination. And the application of a higher standard of law, even after controlling for other relevant factors, increases the probability of a court reaching a disposition favorable to litigants alleging a violation of their rights.

At the same time, though, our study suggests that an ERA will not always be the panacea or the Pandora’s box some occasionally make it out to be. While the results lend support to arguments suggesting that these formal constitutional provisions probably will alter the way courts adjudicate claims of discrimination, our findings also demonstrate that other factors—from the fraction of women comprising the court to the position taken by the government over the suit’s resolution to the facts it entails—likely will impact the efficacy (or lack thereof) of an ERA.

These represent the basic policy implications of our research, but they are in fact, as our emphasis on “probably” and “likely” indicates, qualified implications. That is because we only have been able to assess systematically the quality and certainty of our results with regard to the effect of ERAs in the states. We cannot—in the absence of an amendment to the U.S. Constitution and resulting litigation—make a similar assessment about the federal context. Nonetheless, for reasons we have already mentioned, we believe our analyses provide some insights into the potential effect of a federal ERA—insights that, in general, lend support to Mansbridge’s (1986) reasoned speculation of some fifteen years ago: Even if passage of the ERA leads federal jurists to adopt a strict scrutiny approach to sex-based classification, a step rather likely under our analysis, that approach will not necessarily compel them to strike down the classification as is the case in race litigation.25 Rather, and again as Mansbridge and others (e.g., Case, 2000; Mezey, 2003; Sullivan, 2002) astutely recognized, the outcomes of sex discrimination suits will depend, at least in some part, on the types of judges interpreting the constitutional provision and the particular facts of the suit itself.

24 These probabilities are available on our web site.
25 As we indicated in note 3, affirmative action litigation is a notable exception.
If this is the primary policy contribution of our study, then the primary implication for more
general analyses of legal decisions is clear: Just as so many judicial specialists have suggested, we
cannot and should not count on rules and principles of law to do all the work in explaining the
choices judges make; comprehensive accounts require a consideration of a range of forces. At the
same time, though, we cannot and should not discount institutions, whether formal or informal
in nature. To be sure, as we emphasized earlier, few students of judicial politics think in this
way: at least in theory most believe that rules play an important role in adjudication. But theory
occasionally turns out to be different than practice, as exemplified by the many papers—and recent
ones at that—that omit rules from their analyses of court decisions.

Our study suggests that this omission is not optimal. At minimum, failure to incorporate
institutions amounts to underspecification; at most, it serves to perpetuate a myth about *judging*:
that it is a phenomenon largely (or, in some studies, exclusively) about politics, and not law.
The lesson of our study, as well as of other contemporary analyses, is that judging is about both;
and only by characterizing it as such are we likely to develop more apt descriptions and richer
explanations of how jurists operate.
Appendix A. Description of the Variables (N=416)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>State ERA</td>
<td>.35</td>
<td>.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Court Preferences</td>
<td>44.05</td>
<td>18.88</td>
<td>.96</td>
<td>92.36</td>
</tr>
<tr>
<td>ERA Ratification</td>
<td>.47</td>
<td>.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Female Justices</td>
<td>.06</td>
<td>1.00</td>
<td>0</td>
<td>.43</td>
</tr>
<tr>
<td>Appellate Court</td>
<td>.74</td>
<td>.44</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Standard of Law</td>
<td>1.72</td>
<td>.75</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Physical Difference</td>
<td>.23</td>
<td>.42</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Government Participation</td>
<td>.63</td>
<td>.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Female Litigant</td>
<td>.29</td>
<td>.46</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Outcome</td>
<td>.41</td>
<td>.49</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Appendix B. A Bivariate Mixed Response Probit Model

We are interested in modeling two dependent variables that we observe at the same time. The first variable of interest is called standard, and the second called outcome. We do not assume that these choices are independent, and thus do not fit two separate models. Rather, we explicitly model the dependence with a bivariate mixed response probit model that allows for correlation across two equations.

Here is the notation to be used throughout:

- $i = 1, \ldots, N$ indexes the cases.
- $y_{1,i}$ is the standard chosen in each case. This is an ordinal variable which can take values $1, 2, \ldots, K$. In our application $K = 3$.
- $x_{1,i}$ is a column vector of covariates that explain the standard chosen on the case. It is of dimensionality $(P_1 \times 1)$.
- $y_{2,i}$ is the outcome reached in each case. This is a dichotomous variable, which takes a value of 1 is the party alleging discrimination won the case, and 0 otherwise.
- $x_{2,i}$ is a column vector of covariates that explain the outcome reached in the case. It is of dimensionality $(P_2 \times 1)$.

The model we estimate is essentially an ordinal probit model (McKelvey and Zavoina, 1975) and a dichotomous probit model with correlated errors.\(^{26}\) We assume that our observed dependent

---

\(^{26}\)This particular application is a special case of a bivariate ordinal probit model. The model is easily extended to allow for continuous outcomes, and any number of equations. However, if one were to move beyond two dependent variables, inference using maximum likelihood would become essentially impossible. This is the same problem that plagues estimation of multinomial probit models. Performing Bayesian inference, using Markov chain Monte Carlo methods, would be a promising avenue for estimating higher dimensional models.
variables $y_{1,i}$ and $y_{2,i}$ are governed by latent (unobserved) utilities in the standard fashion, denoted $z_{1,i}^*$ and $z_{1,i}^*$, which are modeled by the covariates $x_{1,i}$ and $x_{2,i}$.

The first equation is the model for \textit{standard}, which is an ordinal variable taking values $1, 2, \ldots, K$. We assume that:

$$z_{1,i}^* = x_{1,i}' \beta_1 + \varepsilon_{1,i}$$  \hspace{1cm} (1)

Where $\beta_1$ is a ($P_1 \times 1$) column vector of parameters to estimate. The probability that $y_{1,i}$ takes particular values is determined by the distributional assumption, the $\beta_1$ parameters, and a set of cut-points $\tau_0, \ldots, \tau_K$. Let $\tau$ denote the set of all cut-points. To identify the model, by construction $\tau_0 = -\infty$, $\tau_1 = 0$, and $\tau_K = \infty$. Before imposing the restrictions, there are ($K + 1$) cut-points; after identification, there remain ($K - 2$) to estimate. Given this specification, the marginal probability of observing a particular standard is:

$$\Pr(y_{1,i} = k) = \Phi(\tau_k, x_{1,i}' \beta_1) - \Phi(\tau_{k-1}, x_{1,i}' \beta_1)$$  \hspace{1cm} (2)

Where $\Phi(\cdot, \mu, \sigma^2)$ denotes the cumulative distribution of the Normal distribution with location parameter $\mu$ and scale parameter $\sigma^2$. The location of the latent utilities with respect to the cut-points is what determines the probability of observing certain outcomes.

The second equation is the model for \textit{outcome}, which is a dichotomous variable. We assume that:

$$z_{2,i}^* = x_{2,i}' \beta_2 + \varepsilon_{2,i}$$  \hspace{1cm} (3)

Where $\beta_2$ is a ($P_2 \times 1$) column vector of parameters to estimate. The marginal probability that the party alleging discrimination lost the case is:

$$\Pr(y_{1,i} = 0) = \Phi(0, x_{2,i}' \beta_2)$$  \hspace{1cm} (4)

The marginal probability that the party alleging discrimination won the case is this probability subtracted from one. A positive latent utility corresponds to a case win; a negative utility corresponds to a case loss.

If we were willing to assume that the \textit{standard} and \textit{outcome} choices were independent, we could assume that $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$ were independent, and distributed according to a standard Normal distribution. Since the scale on which the latent utilities are measured is not identified, one must fix the variance to identify the model. This would result in an ordinal probit model (Equation 1) and a dichotomous probit model (Equation 3), which could be estimated equation-by-equation using maximum likelihood and standard software. In this application, however, it makes no substantive sense to assume that the errors are independent. Indeed, we would expect some correlation among the errors. We thus assume that the errors follow a bivariate Normal distribution:

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \end{bmatrix} \sim \mathcal{N}_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

The diagonal elements of the variance-covariance matrix are normalized to one to identify the scales on which the latent utilities are measured. The parameter $\rho$ is a correlation parameter which gauges the extent to which the errors are correlated. One could stack Equations 1 and 3 to form a seemingly unrelated regression (SUR) model in the latent utility space.

To compute the joint probability of observing a particular outcome, one would need to integrate the bivariate Normal distribution over the correct region. For example:

$$\Pr(y_{1,i} = 3, y_{2,i} = 1) = \int_{\tau_2}^{\tau_3} \int_{0}^{\infty} \phi_2(\mu_i, \Sigma) dz_{2,i} dz_{1,i}$$
Where:
\[
\mu_i = \begin{bmatrix} x'_{1,i}\beta_1 \\ x'_{2,i}\beta_2 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}
\]

And \(\phi\) denotes the probability density function of the bivariate Normal distribution with mean \(\mu_i\) and variance-covariance matrix \(\Sigma\). The probabilities of all other outcomes are computed similarly.

**Estimation**

We estimate the model using maximum likelihood. The log-likelihood function for the model is:

\[
\ln \mathcal{L}(\beta_1, \beta_2, \tau, \rho \mid y_{1,i}, y_{2,i}) = \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbb{I}(y_{1,i} = k) \left\{ y_{2,i} \ln \left[ \int_{\tau_{k-1}}^{\tau_k} \int_{-\infty}^{\infty} \phi_2(\mu_i, \Sigma) dz_{2,i} dz_{1,i} \right] \right. \\
+ (1 - y_{2,i}) \ln \left[ \int_{\tau_{k-1}}^{\tau_k} \int_{-\infty}^{0} \phi_2(\mu_i, \Sigma) dz_{2,i} dz_{1,i} \right] \right\}
\]

Where \(\mathbb{I}(\cdot)\) is an indicator function that takes the value 1 if \(y_{1,i} = k\), and 0 otherwise. We maximize the log-likelihood function using the BFGS algorithm (Nocedal and Wright, 1999, 194-201), using the \texttt{R} (Ihaka and Gentleman 1999) implementation in the \texttt{optim()} function. There are no known analytical methods for computing the rectangular integrals in the likelihood function. We use the simulation method of Genz (1992), which is implemented in the \texttt{sn} library (Azzalini, 2002). For problems of modest size (\(N \approx 450\)), the optimizer takes two to three hours to converge on a dedicated Linux workstation. We use the equation-by-equation maximum likelihood estimates for starting values, which speeds convergence considerably.\(^{27}\) To compute the White (1980) standard errors, we sum the matrix of scores by the clustering variable, and then use their weighted outer product to compute the variance-covariance matrix of the estimator.

**Interpretation**

To interpret results from the estimate statistical model we account for parameter uncertainty by drawing parameter values from their (asymptotic) sampling distribution (see King et al., 2000). The predicted probabilities in Figures 1 and 3 come from Equation 2, and those in Figure 4 from Equation 4. The simulation used to account for parameter uncertainty follows the standard approach.

The algorithm used in Figure 3 is a bit more complicated, as it requires us to draw from the conditional distribution \(f(z_{2,i}^* \mid z_{1,i}^*)\). This distribution is a Normal distribution with mean \(x'_{2,i}\beta_2 + \rho(z_{1,i}^* - x'_{1,i}\beta_1)\) and variance \((1 - \rho^2)\). To generate the predicted probabilities, we:

- Draw a vector of parameters from the (asymptotic) sampling distribution.
- For all standards, draw \(z_{1,i}^*\) from a truncated Normal distribution with mean \(x'_{1,i}\beta_1\) and variance one. For \(y_{1,i} = 1\), the distribution is truncated above at zero. For \(y_{1,i} = 2\), the distribution is truncated below at zero and above at \(\tau_2\). For \(y_{1,i} = 3\), the distribution is truncated below at \(\tau_2\).
- Compute the probability of each standard using the conditional distribution above.

\(^{27}\)We have fit the model with dispersed starting values, and the algorithm always converges to the same estimate.
• Repeat the simulation a large number of times.

It is interesting to note that these predicted probabilities do not look Normally distributed.

Appendix C. Notes for Referees

Material corresponding to footnote 9. We searched the state files in LEXIS on the language of all state equality provisions; and on the terms equal rights amendment, equal protection w/10 gender [or sex, or pregnancy substituted for gender], and gender [or sex, or pregnancy substituted for gender] w/10 discrimination. We also extracted cases from extant literature on the subject, reading all cases cited in all relevant articles (along with sources cited in those articles) produced by searches in LEXIS/NEXIS and HeinOnline on the words equal rights amendment, gender w/ 10 discrimination, sex w/10 discrimination, and E.R.A. in the title. Particularly useful sources were: Altschuler (1983, 1992); Avner (1984); Beck (1977); Beck and Baker (1993); Brown et al. (1971); Crump (1973); Driscoll and Rouse (1977); Dybwad (1974); Farone (2000); Gammie (1989); Ginsburg (1975); Gokel (1978); Harvard Law Review (1970); Kanowitz (1967, 1973); Kaufman (2001); Kruger (1986); Kurtz (1977); Linton (1997); McCausland (1983); Minnesota Law Review (1973); Hirczy de Mino (1997); Northwestern University Law Review (1971); Saucier (1980); Schoen (1978, 1983); Sherwin (1984-85); Simpson (1977); Tarr and Porter (1982); Treadwell and Page (1977); Nunziato (1994); Voller (1974); Williams (1985, 1994); Women’s Rights Law Reporter (1974). We also inspected leading texts and case books with the word gender, sex, or women in the title (see, e.g., Babcock et al., 1975; Brown et al., 1977; Goldstein, 1994; Greenberg et al., 1998; Kay, 1981).

Material corresponding to footnote 10. We “shepardized” the same three cases as Gryski et al. (1986): (Reed v. Reed, 1971; Frontiero v. Richardson, 1973; Craig v. Boren, 1976), as well as seven decided prior to the time frame of their study (Bradwell v. Illinois, 1873; Minor v. Happersett, 1875; Muller v. Oregon, 1908; Adkins v. Childrens Hospital, 1923; West Coast Hotel v. Parrish, 1937; Goesaert v. Cleary, 1948; Hoyt v. Florida, 1961) and four subsequent to the Gryski et al. cases (Orr v. Orr, 1979; Mississippi University for Women v. Hogan, 1982; J.E.B. v. Alabama, 1994; United States v. Virginia, 1996) (all of which, for varying reasons, may have particularly important implications for state court litigation).

Material corresponding to footnote 14. Most scholarly lists of “ERA states” identify 18: Alaska (1972), California (1879), Colorado (1973), Connecticut (1974), Hawaii (1972), Illinois (1971), Louisiana (1974), Maryland (1972), Massachusetts (1976), Montana (1973), New Hampshire (1974), New Mexico (1973), Pennsylvania (1971), Texas (1972), Utah (1896), Virginia (1971), Washington (1972), and Wyoming (1890). But there are two complicating factors. The first is that controversy exists over whether New Jersey does or does not have an ERA (see Williams, 1994). A second concern is that while 15 of the 18 states (or 15 of 19 if we include New Jersey) added their ERAs in the 1970s, three are of a much older vintage. In light of our objective of making inferences about the potential effect of a national ERA, we questioned whether to exclude the three. But because they appear on virtually all ERA lists, we decided to code them as such for our primary analyses. We also decided, following the consensus in the literature, against including New Jersey as an ERA state. Of course we assessed the sensitivity of our results to these decisions. The results of that assessment, which are housed on our web site, were remarkably stable and consistent with those depicted in Table 2.
Material corresponding to footnote 15. In the intermediate scrutiny category we also included a variation on that standard—that the government must offer an “exceedingly persuasive justification” for discriminating on the basis of sex—which some U.S. Supreme Court justices have endorsed (see, e.g., United States v. Virginia, 1996) (see also note 4). In the strict scrutiny category, we also included a standard invoked occasionally by a few state courts—a standard that some observers liken to strict scrutiny, while others describe as “stricter” than strict scrutiny because it supposedly does not allow for sex-based classifications. To the extent that courts qualify this “stricter” standard with terms such as “absent compelling justifications” or in cases based on “actual differences,” however, the prohibition is not absolute (thereby lending credence to the view that it is akin to strict scrutiny). For our purposes, though, the key point is that it is closer, if not identical, to strict scrutiny than it is to the intermediate or rational basis standard.

More generally, we should mention that [the paper’s first author] coded all the variables after extensive pre-testing of the coding scheme. We draw (via SPSS) an ≈ 10% sample of the cases coded by [the first author] and compared the results to coding performed by a J.D. student. The coders disagreed over the standards used in three cases, for an error rate of 8.57%. In all three instances, the disagreement was over the next-closest category (e.g., a disagreement over standard=1 versus standard=2). For the full results, see our web site.

Material corresponding to footnote 21. As part of this assessment we incorporated the 51 cases in which the court did not articulate a standard and added the value of 0 to the standard variable to account for these cases; redefined value 1 of the ERA variable to include only states ratifying their amendment in the 1970s and also to include New Jersey; included a variable to differentiate criminal and civil cases; and added the variable representing the proportion of women to the outcome equation. The results, which are housed on our web site, were remarkably stable and consistent with those depicted in Table 2—with one notable exception: though including the proportion of women variable does not alter any of the other results in Table 2, it fails to affect significantly case outcomes.

Material corresponding to footnote 22. The replication archive contains R code to replicate all of the analyses in this paper. This includes the dataset, the results in Table 2, all code used for model interpretation, and all robustness analyses (including robustness to starting values, model specification, and the code used to compute the robust standard errors).
References


