Scandinavian Political Studies
A journal published for the Nordic Political Science Association

ISSN 0080-6757 Doi: 10.1111/1467-9477.12176

The Government Deference Dimension of Judicial Decision Making: Evidence from the Supreme Court of Norway

Jon Kåre Skiple* , Henrik Littleré Bentsen† and Chris Hanretty††

Past research has revealed conflicting findings regarding the degree to which judges on European apex courts enact their policy preferences or instead disagree on the basis of divergent legal views. We investigate disagreement between judges on the Norwegian Supreme Court between 1996 and 2016. During this period, the court dealt with a greater volume of policy-relevant cases than previously. The method of appointment to the court was also changed to a judicial appointments commission. We analyse non-unanimous cases using item response theory models. We find that judges are not divided along left–right lines but instead disagree about the appropriate degree of deference to give to public authorities. There is no significant association between the appointing government and judges' ideal points either before or after the reform to appointments. Judges who were formerly academics are however much less deferential than career judges or judges who were previously lawyers in private practice.

Introduction

Judges on European supreme and constitutional courts have become increasingly influential decision makers over the past thirty years. Their influence has grown thanks to domestic and international factors. Domestically, courts have acquired greater power over their caseload, and this has led to dockets that consist of fewer but more policy-relevant cases. Courts have also been empowered by developments in constitutional and international law relating to the European Convention of Human Rights (ECHR) and the European Union, which increasingly dominate the issue agenda of European high courts (Keller & Sweet 2008; Alter 2010), giving justices new

* Jon Kåre Skiple Department of Comparative Politics, University of Bergen, Postboks 7800, 5020 Bergen, Norway. Email: josk@norceresearch.no.
† Henrik Littleré Bentsen, Department of Comparative Politics, University of Bergen.
†† Chris Hanretty, Department of Politics, International Relations and Philosophy, Royal Holloway, University of London.

opportunities to oversee public authorities (Shapiro & Stone 1994; Shapiro & Sweet 2002; Hirschl 2009; Martinsen 2011). The increased influence of these courts has drawn attention to the methods used to appoint judges to these courts, and to the diversity of justices' socio-demographic and pre-appointment judicial careers (Shaffer et al. 2015).

The growth in the influence of European courts has led to a burgeoning academic literature. This literature has been substantially aided by systematic attempts at data gathering for different apex courts (in Germany, through the DFG-funded project 'The Federal Constitutional Court as a veto player'; in Norway, through work reported in Grendstad et al. (2015b); and in Israel, through the Israeli Supreme Court Database), and by the use of advanced techniques for the analysis of judicial behaviour (Martin & Quinn 2002; Clinton et al. 2004). These developments have led to multiple discoveries suggesting that judges' preferences are more complicated than purely attitudinal or legal models suggest (Epstein & Knight 2013; Hanretty 2015a; Arvind & Stirton 2016). For many countries, we do not yet know what drives disagreement between judges. We are therefore very far from being able to generalize about how institutions structure judicial behaviour across western Europe.

This paper contributes to this literature by developing a new account of judicial disagreement in Norway. In this paper, we make three claims concerning judicial behaviour on the Supreme Court of Norway. First, we show that most disagreement on the court is disagreement about the appropriate amount of deference to show to government, rather than disagreement between left and right. Second, we show that judges' views on deference are structured by their pre-appointment careers: judges who were previously academics are less likely to defer to government than judges who were previously in private practice or in other judicial roles. Third, we show that judges' views on deference are not associated with the identity of the appointing government. A 2002 reform of the appointment process – which moved from ministerial appointment to appointment by independent commissions – therefore mattered more for the perception of independence than actual behaviour on the court. We base these three claims on the analysis of 644 non-unanimous cases decided by the court over a twenty-year span.

Our paper is structured in seven parts. In this first introductory section, we identify the principal literature to which we contribute, and the ongoing issues in that literature. In the section that follows ('Context'), we describe the institution that we analyse, the Norwegian Supreme Court, providing an overview of its structure and operation. We also give reasons why the Norwegian Supreme Court is an interesting court to study. In a third section ('Literature'), we discuss two main issues – the nature of judicial preferences and the link between appointment and behaviour – in the light of comparative and Norwegian debates. We describe the data that we use before

moving on to describe the method (item response theory) used to analyse those data. Our sixth section presents our results as they relate to individual judges and groups of cases. In our conclusion, we draw three implications for the broader study of judicial behaviour and institutional design.

Context

The Norwegian Supreme Court

The Norwegian Supreme Court was established in 1815 following the adoption of the Norwegian Constitution in the previous year. The court functions exclusively as an appellate court. It has no original jurisdiction, and matters of constitutional interpretation are dealt with only in order to resolve appeals before the court. The court ordinarily sits in panels of five justices. In more important matters, cases may be heard in a grand chamber (11 judges) or in a plenary session (19 judges). Since 1863, judges have had the opportunity to publicly express their individual opinions through dissents and concurrences (Mestad 2015). Because Norway's legal culture mixes the civil law system's reliance upon codes, laws and statutes with the common law system's reliance on judges and precedents (von Eyben 1956; Sunde 2015), Norwegian justices have the opportunity to take a more active role in policy making and the development of law than most of their civil law counterparts (Skiple et al. 2016).

As a generalist apex court that decides cases concerning constitutional issues as part of a broader caseload, the Norwegian court is more similar to apex courts in common law legal systems than specialized constitutional courts in Central and Eastern Europe. Because the court has the competence to review the legality of an administrative action and to determine the constitutionality of Acts of Parliament (i.e., judicial review), the court has an important position in the political system (Magnussen 2005).

Appointments to the Court

Until 2002, the judges of the Supreme Court were appointed by the King-in-Council, acting on the recommendation of the Ministry of Justice. This system was similar to the systems of ministerial appointment found in several common law countries (Hanretty 2015b). In 2001, the Norwegian Parliament approved reforms to the court system which also changed the system for appointing judges to the Supreme Court. Since 2002, judges have been appointed by the King-in-Council acting upon the recommendation of a seven-member Judicial Appointments Board. The Judicial Appointments Board must be made up of 'three judges, one lawyer, one lawyer in public service, and two non-lawyers' (Grendstad et al. 2015b, 47). The Appointments

Board announces vacancies, vets applicants and makes a final recommendation by ranking the top candidates (Sunde 2015). No government has failed to follow the Board's recommendations. Because justices serve until a mandatory retirement age of 70, appointments to the court are naturally staggered, and our data include votes from justices appointed substantially before and after the 2002 reform.

Cases Heard by the Court

As a general appellate court, the court hears a variety of cases. Its caseload in the period studied here (1996–2016) is, however, quite different to its caseload earlier in the court's history. In 1995, a reform of the criminal procedures relieved the Court of the heavy burden of serving as a default court of appeals for minor criminal cases, leading to a significantly lighter caseload (Sunde 2015). This coincided with gradual institutional reform in civil procedures. The combined effect was to grant the court near-complete discretionary power over its docket. These changes lead to a lower caseload composed of more policy-consequential cases, which, in turn, was followed by an increase in dissents on the court (Bentsen 2018). More broadly, the reform enabled the court to concentrate on its agenda around appeals that raise issues of importance beyond the present case to fulfil its function as a court of precedent that – through its caselaw – works towards clarification and development of law. Thus, the litmus test for granting leave to appeal is whether there is 'something to gain for the legal system and for the society as a whole' (Bårdsen 2018, 11).

A fair share of cases on the court deals with constitutional issues, either as questions concerning judicial review of acts of parliament (decided in plenary), reviewing the constitutionality of administrative actions or using constitutional arguments to interpret and apply statutory law (Bårdsen 2017). The court very rarely exercises judicial review to its fullest effect (i.e., finding a law unconstitutional); however, in 2010 the court set aside laws enacted by the parliament in three cases, thereby demonstrating the court's power to perform judicial review. In the plenary decision in the Ship Owner Taxation case (casenumber: Rt-2010-143), the court held – by a majority of six to five – that a tax law enacted by the parliament violated the prohibition against retroactive legislation in Article 97 of the Norwegian Constitution, consequently barring the government from collecting a potential tax income of 21 billion NOK (USD 3.547 bn) (Grendstad et al. 2015b, 166).

The Supreme Court and International Courts

The Supreme Court coexists with two international juridical regimes. First, following the 1994 ratification of the European Economic Area (EEA) agreement, the Supreme Court must follow decisions of the Court of Justice of the

European Free Trade Association States (the EFTA Court) in matters relating to the European single market, and may refer issues to the court where the relevant provisions are unclear. Because the single-market legislation is EU legislation, and because the EFTA court follows the jurisprudence of the Court of Justice of the European Union (CJEU), this means that the Supreme Court also must have awareness of developments in that court. Second, the Human Rights Act of 1999 incorporated the provisions of the European Convention on Human Rights into domestic Norwegian law, in practice obliging the Supreme Court to heed the jurisprudence of the European Court of Human Rights (ECtHR). In 2014, the constitution was amended to include a new chapter on human rights. This raises the possibility that the court could develop its own autochtonous human rights jurisprudence. Generally, developments since the 1990s have meant that international law has become a substantial part of the agenda of the court, giving justices new opportunities to exercise control over public authorities (Wiklund 2008; Martinsen 2011).

Justifying the Study of the Court

Having described the court and its operation, we can now better justify our decision to analyse judicial behaviour on the court. Four characteristics of the court make it an exciting case to study. First, by identifying and characterizing the bases of judicial disagreement on a generalist apex court, we add to the growing literature on apex courts in Europe, which have mainly been concerned with specialized constitutional courts (e.g., Hönnige 2009; Pellegrina & Garoupa 2013; Coroado et al. 2017; Pellegrina et al. 2017; De Jaegere 2019; Popelier & Bielen 2019). Second, the change in the court's docket means that policy-relevant cases are much more common. This in turn makes it more likely that policy preferences will emerge in a distinct form (Hall 1985; Brace et al. 2012). Alternately, if policy preferences do not emerge, it will not be for a lack of opportunity. Third, the presence of international agreements allows for a specific type of policy preference to emerge, namely a centre–periphery domestic/supra-national cleavage (Malecki 2012; Voeten 2007). Fourth, and perhaps most importantly, the change in the appointment mechanism allows for a test of the effect of different institutions upon judges' preferences. We describe some of the expectations relating to appointment institutions in the section which follows.

Literature

The Nature of Judicial Preferences

It is common in the study of judicial behaviour to explain what judges do by pointing to their preferences (Epstein & Knight 2013). Although the range

of possible preferences is broad, many influential studies have focused on judges' policy preferences on the same left–right scale that structures party politics in many countries. Judges may either straightforwardly enact those preferences (the attitudinal model: Segal & Spaeth 2002) or may be constrained (the strategic model: Bergara et al. 2003; Brouard & Hönnige 2017). These preferences are either inferred from patterns of judicial (dis) agreement, or they are inferred from the position of the actor appointing each judge. These kinds of inferences are found whenever appointing party is used as a proxy for judicial preferences. Previous studies which have inferred judges' preferences on a left–right scale have included multiple studies of the US Supreme Court (Martin & Quinn 2002) and American state supreme courts (Windett et al. 2015), as well as studies of European constitutional courts (Hanretty 2012, 2014). This focus on left/right preferences in political science is understandable: the left/right spectrum structures most of electoral politics in advanced industrial democracies, and political scientists who are interested in courts tend to be most interested in 'politicised' courts, which are more likely to reflect the same preference structures found in electoral politics.

Preferences beyond Left and Right

There is, however, an emerging literature on other kinds of preferences which, although they are political, are less obviously connected to the main dimensions of political conflict. Scholars of judicial politics in western Europe ought to be more attuned to these possibilities, since two regional courts in the CJEU and the ECtHR demonstrate judicial preferences that are not left/right preferences but are closer to centre/periphery preferences (Malecki 2012; Voeten 2007). In other (national) courts, there is evidence that, although judges' preferences can be distinguished, these preferences relate to the appropriate degree of deference to give to public authorities – preferences which are partly political (insofar as they relate to the appropriate degree of state intervention) and partly legal (Arvind & Stirton 2016) – or to the appropriate degree of support to give to plaintiffs rather than respondents (Hanretty 2015a). It is not yet clear whether apex courts in Western Europe more closely follow the model of the Supreme Court of the United States, where left/right preferences are paramount, or whether instead they are characterized by subtler political differences.

Appointments and Preferences

The literature on judicial preferences has little in common with the literature on appointment methods. This literature is most strongly related to the theme of judicial independence, although it has occasionally been hard to link merit-based or depoliticized appointment mechanisms (a component

of most measures of de jure judicial independence) to de facto judicial independence (Melton & Ginsburg 2014). There is good evidence that different systems lead to judges with different observable characteristics: Valdini and Shortell (2016) show that 'exposed' appointment mechanisms (where appointments are made by publicly accountable actors such as elected politicians) are associated with a greater proportion of female appointees than 'sheltered' appointment mechanisms (where appointments are made by merit-based appointments commissions). Insofar as judicial characteristics can affect preferences (Tate 1981), this might be the basis for a link. The closest to a direct test of the link between appointment mechanisms and the structure of judicial preferences is Weinshall et al. (2018), who, in a study of decision making on five apex courts in the US, Canada, Israel, India and the Philippines find some support for the claim that 'when the appointment process involves other players in addition to the political branches of government – for instance in the form of judicial appointment committees consisting of both politicians and legal professionals – appointments… yield a less attitudinal bench' (Weinshall et al. 2018, 341; see also Carroll & Tiede 2011).

The Norwegian Debate

The preceding paragraphs have discussed the English language social scientific literature on judicial behaviour. In the country we study, there is an ample debate about how best to characterize decision making on the Norwegian Supreme Court, and regarding the effects of changes to methods of appointment. Part of this discussion has involved the real effect of the 2002 reform to the method of appointment. The intention of the reform was to eliminate any politicized appointments to the court. Statements by former ministers of justice on appointments, as well as de facto appointments of former ministers of justice to the Court demonstrate that well before 2002, the government could have used the appointment process to place ideologically like-minded justices on the Court (Grendstad et al. 2015b, ch. 2). In contrast, through interviews and historical data, Sunde (2015, ch. 12) argues that the appointment process before 2002 was not influenced by ideological considerations. Rather, Sunde argues, the process reflects a modest degree of self-recruitment by the Supreme Court on the basis of candidates' legal credentials. The Chief Justice meets the Minister of justice on a regular basis, and the Chief is always permitted to offer views on the applicants before their final appointment.

There is a partly related debate about the background of justices, and the effects of these personal attributes on judges' behaviour. Several scholars have argued that different backgrounds affect judges' deference. One dominant position holds that justices with pre-appointment careers in

government administration are more inclined than other justices to defer to the public party (Lund 1987; Kjønstad 1997, 1999). Some empirical studies bear this out (Grendstad et al. 2015a, 2015b; Skiple 2015; Skiple et al. 2016). A second position holds that academics are more likely to challenge the government. Studies of the Norwegian Supreme Court identify a number of legal academics on the Supreme Court as being among those expressing a stronger desire to 'make law'. They actively use the constitution as a legal force and vigorously pursue Norway's commitment to international law and human rights (Kinander 2016; Robberstad 2016), and are generally more prone to disagreement (Bentsen 2018).

Data

Scope

We use data from all non-unanimous cases decided by the Supreme Court of Norway between January 1996 and December 2015 in panels of five, 11 or 19 judges. We only use non-unanimous cases because (without very strong assumptions) unanimous cases are uninformative about the nature of disagreement on the court in item response theory models (Martin & Quinn 2002, 137fn3). We use decisions after 1996 because this was the first full year following the reforms to the court's caseload which eliminated a large number of routine criminal appeals. We use decisions until 2016 because this is the last year for which we have full data. In total, we analyse data from 644 cases decided by 47 judges. We omit judges who sat on these cases but who decided fewer than 10 cases. The resulting total number of justice votes is 3,458.

Data Collection and Coding

We use data from Norwegian Supreme Court Database (Grendstad et al. 2015b). Decisions are manually collected by the NSCD team from the judicial text database Lovdata.no. Judge appointments, careers and background are collected from the National Archives of Norway and historical and contemporary sources. All variables in NSCD have been coded according to predefined codebooks. From NSCD we gathered variables on judges' votes, appointments and careers, case parties, case winner, issue type and panel size from 1996 to 2016. In addition, we complemented NSCD data with three variables: constitutional cases, criminal sentencing cases and judges' main careers. To identify (1) constitutional and (2) criminal sentencing cases, we (1) searched decisions for legal references to the Constitution ('Grunnloven') and (2) searched the keyword section of the decision for criminal sentencing ('straffutmåling') and community sentencing ('samfunnsstraff'). We did

both searches on a text corpus of the decisions extracted from Lovdata.no using *domstolr* (Bjørnebekk & Johannesson 2016). We collected data on main pre-appointment career from judges' official bios on the supreme court's webpage and the Norwegian Encyclopedia. We counted the number of years each judge had worked in either government administration, private law firms, lower courts or academia. The occupation with highest number of years was coded as the judge's main occupation.

The Dependent Variable

The dependent variable in our analysis is whether or not the relevant judge voted with the majority or dissented. By a 'dissent' we include both disagreement with the outcome and disagreement in reasoning. Around 30 percent of the 3,458 votes are dissents, and the majority of cases only feature one dissenting vote. By convention we record votes with the majority as having a value of one, and dissenting votes as having a value of zero, but nothing depends on this decision.

Case-level Variables

We use eight case-level variables. These are as follows:

- whether the outcome in a civil case, meaning the broader category of non-criminal cases, was favourable (102 cases) or unfavourable (86) for the public authority, or neither (both sides were private or both sides were public);
- whether the outcome in a case involving the European Convention on Human Rights was favourable (48 cases) or unfavourable (44) for the public authority, or neither (both sides were private or both sides were public);
- whether the outcome in a case involving a reference to European Union (EU) law or directives to case law from the European Court of Justice (ECJ) was favourable (9 cases) or unfavourable (18) for the public authority, or neither (both sides were private or both sides were public);
- whether the outcome in a criminal case was favourable (109 cases) for the prosecutor or for the defendant (131);
- whether the outcome in a criminal sentencing adjustment case was favourable (65 cases) for the prosecutor or for the defendant (76);
- whether or not the outcome in a case involving a reference to the constitution was favourable (23 cases) or unfavourable (21) for the public authority, or neither (both sides were private or both sides were public);
- whether or not the outcome in a case involving economic matters was favourable (52 cases) or unfavourable (53) for the public authority;

- whether or not the outcome in a case heard by a larger (11 judges or more) panel was favourable (9 cases) or unfavourable (14) for the public authority, or neither (both sides were private or both sides were public).

Either of these variables is scored 1 if the outcome was favourable to the public authority (or prosecutor), –1 if the outcome was unfavourable for the public authority; and 0 if the outcome does not apply (for example the variable asks about outcomes in civil cases, but the case is a criminal law case), if neither side was a public authority, or if both sides were public authorities. By outcome favourable to public authority (or prosecutor), we mean any case where a public authority (or prosecutor) appealed, where the respondent was not a public authority (or prosecutor), and where the appeal was allowed, or alternately where a public authority (or prosecutor) was the respondent, where the appellant was not a public authority (or prosecutor), and where the appeal was dismissed.

We included these eight case-level variables as they offer a basis of what may drive divisions on the court. A deference division in civil cases and constitutional issues (e.g., Grendstad et al. 2015b), a globalist/nationalist division in cases concerning ECHR and EU/EEA (e.g., Malecki 2012), a left–right division in economic cases (e.g., Grendstad et al. 2015b; Skiple et al. 2016) and a soft-on-crime versus hard-on-crime in criminal law and criminal sentencing cases (e.g., Grendstad et al. 2015b). We include cases decided in enlarged panels to account for the possibility that the more principled or consequential cases decided in these panels activate a higher degree of division between the justices.

Judge-level Variables

We include three variables which relate to judges. These are as follows:

- whether or not the judge was appointed by a social democratic government (code: 1; 30 judges) or a conservative government (code: 0; 17 judges);

- whether or not the judge was appointed after (code: 1; 19 judges) the 2002 reform of the appointment system, or before (code: 0, 28 judges); and
- whether the judge's primary pre-appointment career was spent in government administration (23 judges), as a private attorney (10), as a lower-court judge (5), or as an academic (9 judges; the reference category).

In the Appendix we also show models which include information on the judge's gender (14 of 47 judges were female); whether or not they were born in Oslo (22 were); whether or not they served as a temporary judge; and (for

judges previously in government administration) whether they worked in the legislation department, in the Attorney General's office or as a public prosecutor.

Method

Item Response Theory

We analyse judges' votes using a hierarchical item response theory (IRT) model. IRT models are a common way of analysing votes from deliberative bodies like legislatures and courts. IRT models attempt to find judge positions (along some dimension) that best explain the observed pattern of votes. Ordinarily, judges who agree with each other more are more likely to have positions which are closer together on this 'recovered dimension', and judges who disagree more with each other are more likely to be further apart. IRT models, however, use more information than just rates of agreement, because they also allow for the relationship between each case and the recovered dimension to vary: some cases will have high absolute values of a 'case discrimination parameter' (analogous to a high factor loading in factor analysis). These highly discriminating cases will provide an insight into the nature of the recovered dimension, which is not fixed by the model but which must, once recovered, be interpreted by the analyst.

IRT and Testing Expectations

IRT models are compact summaries of patterns on the court under a particular model of decision making. As such, they are primarily descriptive rather than hypothetico-deductive. The language used in the previous section, describing independent and dependent variables, is therefore partly misleading. It is, however, possible to incorporate tests of relationships within IRT models, by modelling some of the parameters in the model as functions of other covariates. In this way, it becomes possible to test the (judge-level) relationship between appointing governments and judge ideal points; or to test the (case-level) relationship between the direction of a public law case (for or against the government) and the sign of the discrimination parameter.

The model can be specified as follows. Let y_{ij} stand for the vote of judge j (1,...,47) in case i (1,...,644), where $y_{ij}=1$ when the judge votes with the majority, and $y_{ij}=0$ when the judge dissents. The probability of the judge joining with the majority is modelled as a function of a judge ideal point (θ_j), a case discrimination parameter (β_i) and a location parameter (α_i):

$$logit\left(Pr\left(y_{ij}=1\right)\right)=\beta_i\theta_j-\alpha_i$$

For cases with a positive discrimination parameter ($\beta_i > 0$), judges who have higher values of θ (either 'more' of the latent trait, or judges positioned on the right-hand or easterly side of the recovered dimension) will be more likely to vote with the majority. For cases with a negative discrimination parameter, it is judges with lower values of θ who will be more likely to end up in the majority.

Ours is a hierarchical IRT Model

This means that the case discrimination parameters are in turn modelled as the product of different covariates and associated coefficients. Specifically, we model

$$\beta_i \sim N\left(X_i\gamma, \sigma\right)$$

where X is a matrix containing an intercept and the eight case-level independent variables listed above. Modelling the case discrimination parameters in this way has three advantages. First, it makes our estimates more precise because we are able to exploit extra information about our cases. Second, it helps in the interpretation of the recovered dimension. If the γ coefficients are all positive and distinct from zero, then it means that pro-public outcomes are associated with higher values of the latent trait – or alternately, that the recovered dimension involves deference to public authorities. Third, it helps us better ground comparisons across time (Ho & Quinn 2010, 845). Some of the judges in our sample never served together, and so we do not have direct evidence of their rates of agreement. For any pair of judges, we can identify a bridging judge or judges who served with both. However, if the set of cases decided by those sets of judges changes, these bridges can be poor bridges. By linking the case discrimination parameters to outcomes, we can be more confident about how judge positions relate to observable case outcomes. This does not solve the problem of intertemporal comparisons – the meaning and nature of ruling in a pro-Convention right direction can also change over time – but it does reduce it.

In common with most researchers using IRT models in political science, we estimate this model using Bayesian methods, and specifically using Stan (Stan Development Team 2018). We use different weakly informative prior distributions for our parameters. The location parameters (α) are drawn from a normal distribution with a mean of zero and a standard deviation of 10. The coefficients on the case-level covariates (γ) are drawn from a Cauchy distribution with a location parameter set to zero and a scale parameter set to 10 for the intercept and 2.5 for all other parameters (Gelman et al. 2008). The residual standard deviation on the case discrimination parameters (σ) is giving an improper uniform prior on the interval $(0,\infty)$. The judge ideal

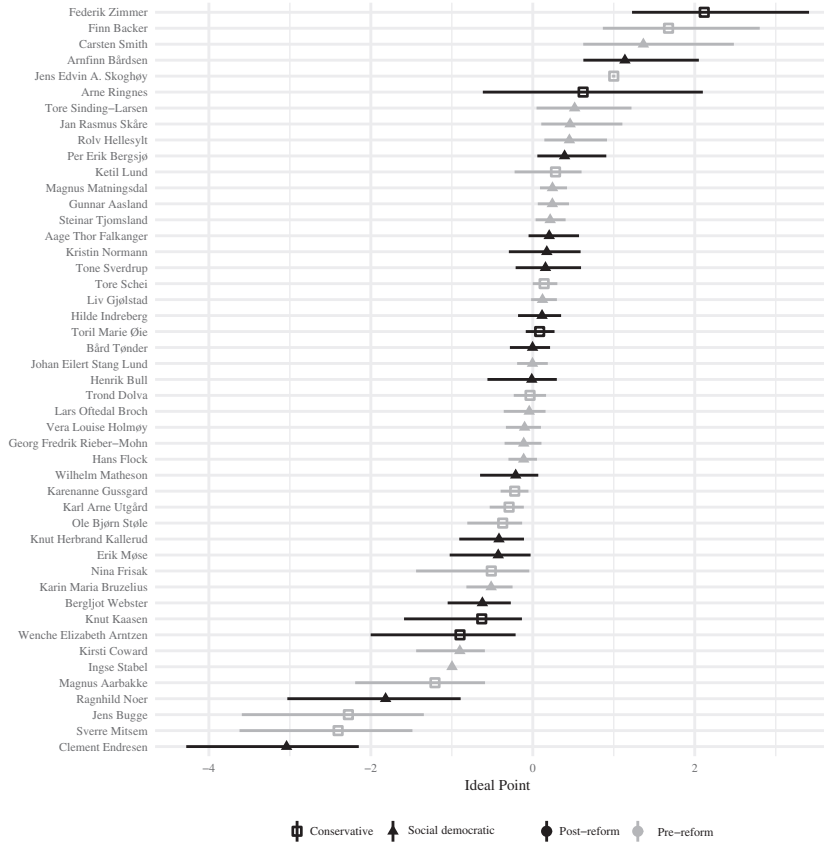
points are drawn from a normal distribution with mean zero and a standard deviation of one. To identify the model, we set one justice to have a positive idea point and one justice to have a negative ideal point. We set Justice Jens Edvin Skoghøy to have a positive ideal point, as he is identified as being more defiant towards the elected branches of government, and set Justice Ingse Stabel to have a negative ideal point because she is identified as being more deferential to the elected branches of government (Grendstad et al. 2015b, 11). Both Justice Skoghøy (1998–2016) and Stabel (2001–2017) served on the court for the majority of the period analysed in this paper. They also represent two different judicial pre-appointment careers and were appointed by different governments. Whereas Justice Skoghøy spent most of his pre-appointment career in the university as a law professor and was appointed by a conservative government, Justice Stabel spent most of her pre-appointment career working in government administration and was appointed by a social democratic government.

Results

The principal output of an IRT analysis of voting in political institutions is a graph which shows the relative positions of the actors casting those votes. This graph helps make sense of the recovered dimension. Before that graph can be shown, however, it is necessary to know whether the IRT model provides a good fit to the data – or at least, a better fit than a reasonable null model. If the model does not provide a good or acceptable fit to the data, then little store can be set by the ideal points recovered by that model. Two measures of fit are the percentage of decisions correctly predicted (PCP) and the geometric mean probability of model predictions (GMP). Both of these measures run from zero to one, where higher values indicate better fit. PCP is the simpler measure, but unlike GMP it does not take into account whether correct predictions were assigned a very high probability or a probability only just greater than 50 percent. Fortunately, both measures of fit agree that the IRT model outperforms the null model. The PCP for the IRT model is 74.7 percent. This compares favourably to the PCP of a null model which simply states that each judge always votes with the majority. This null model has a PCP of 69.4 percent (which is simply the proportion of votes cast with the majority). The GMP of the IRT model is 0.705, which compares to the GMP of a null model which states that each judge probabilistically votes with the majority with probability 0.694. The GMP for this null is 0.54.

Having established that the fit of the IRT model improves on the fit of a null model, we can move on to graphing the judges' positions on the recovered dimension. These positions are shown in Figure 1, together with 90 percent credible intervals surrounding each point estimate. The two judges

Figure 1. Judge ideal points according to appointing government and appointment before or after the 2002 reform. Plotted points give posterior means; lines show 90 percent credible intervals.



chosen as anchor judges (Skoghøy and Stabel) are located far apart towards each end of the recovered dimension. Since these judges have previously been identified as being activist and deferential respectively, this gives initial support for the claim that the recovered dimension separates deferential from less-deferential judges. However, they are not the most extreme judges: judge Endresen is clearly further to the 'left' than judge Stabel, and judge Zimmer is clearly to the 'right' of judge Skoghøy. The ideal points are relatively precisely estimated, which means it is possible to distinguish clearly between positions even of judges in the middle of the spectrum (say, between the position of Tore Schei and Karenanne Gussgard).

In order to interpret better the recovered dimension, we next look at the coefficients associated with the case discrimination parameters. These are

plotted in Figure 2. Where the coefficient is positive, cases with this feature are more likely to feature judges from the right-hand side of the recovered dimension (e.g., judges Skoghøy, Carsten Smith and Zimmer) in the majority; where the coefficient is negative, cases with this feature are more likely to feature judges from the opposite side of the recovered dimension.

There is a strong negative association, clearly distinguishable from zero, between pro-public findings in civil cases and the discrimination parameter. This association is similar in size to the association between pro-public findings in constitutional cases – although this association is less precisely estimated, and as such not significantly different from zero. This suggests that judges on the left-hand side of the recovered dimension (Stabel; Clement Endresen) are more likely to be deferential – that is deliver rulings in favour of public authorities.

We argue that the recovered dimension is associated with deference to public authorities, rather than a left/right or centre–periphery cleavage. The coefficients associated with pro-public authority rulings in cases relating to economic matters (which Grendstad et al. (2015b) take to be indication of left/right preferences) are not significantly different from zero. Nor are the coefficients associated with ECHR or EEA matters. Indeed, these coefficients have different signs, which is not what one would expect if the recovered dimension separated 'localist' from 'globalist' judges.

Figure 2. Coefficients on case characteristics. Lines indicate 95 percent credible intervals.

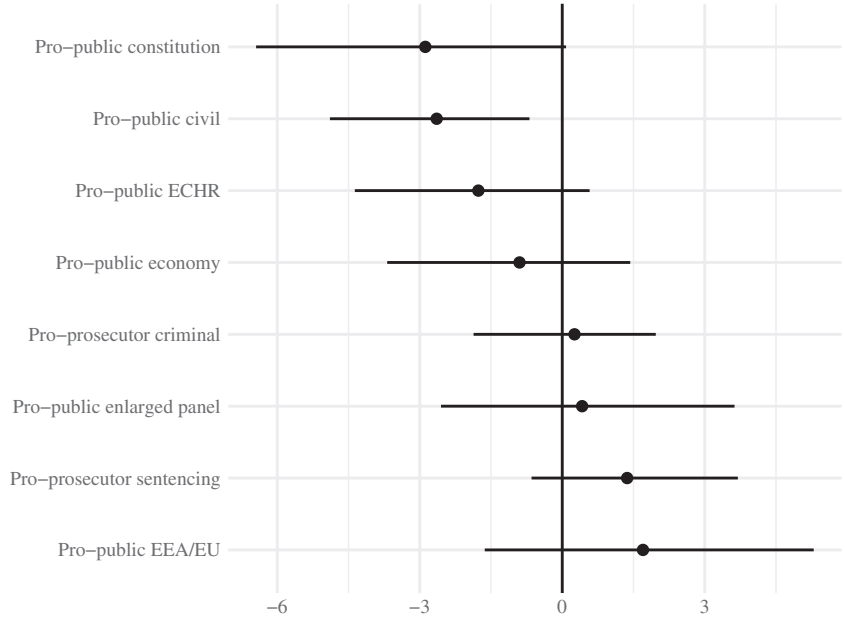


Table 1. Ordinary Least Squares Regression Models of Judicial Ideal Points as a Function of Appointing Actor (Models 1, 2) and Previous Career (Model 3)

	Model 1	Model 2	Model 3
Appointed under Soc-Dem govt	0.063	0.404	
	(0.300)	(0.366)	
Appointed after 2002 reform	–0.051	0.611	
	(0.294)	(0.510)	
Soc.-Dem appointee post-2002 reform		–0.975	
		(0.619)	
Formerly lawyer in private practice			–1.036*
			(0.418)
Former civil servant			–0.586
			(0.357)
Career judge			–1.077*
			(0.507)
Num. Obs.	47	47	47
R^2	0.001	0.056	0.150
Adj. R^2	–0.044	–0.010	0.091
AIC	135.8	135.1	130.2
BIC	143.2	144.4	139.5
Log. Lik.	–63.892	–62.571	–60.111

* $p < 0.05$; ** $p < 0.01$.

Having identified the recovered dimension as a 'deference' dimension, we can now move on to look at the effect of appointments, and the change in appointment mechanism, on judges' ideal points (Table 1). Generally, there is no relationship between the identity of the appointing government and judges' ideal points: the average ideal point of judges appointed by Social Democratic led governments is –0.12 (SD = 0.82), compared to –0.17 (SD = 1.18) for all other governments, a difference which is not statistically significant. Nor is there any difference between pre- and post-reform appointments: pre-reform appointments have an average ideal point of –0.16 (SD = .89) compared to a figure of –0.16 for post-reform appointments (SD = 1.07). Finally, there is no evidence that the relationship between the identity of the appointing government and judge ideal points changed before and after the reform: a regression model which includes an interaction between appointing government and reform shows no significant effects, and the model in total explains almost no variation in judicial ideal points.

We are, however, able to explain part of the variation in judicial ideal points by pointing to judicial backgrounds. Relative to the reference category of academics, judges with a previous judicial background, or a previous background in private legal practice, have ideal points that are much more negative – which means much more deferential. Phrased differently, academics are much more likely to rule against the government than career judges or lawyers. Surprisingly, individuals who made their career within

government (typically the Ministry of Justice), and who might have been expected to have been more deferential, cannot be distinguished on average from academics. These factors explain around a quarter of the variation that we see in judicial ideal points.

Conclusions

In this article we have analysed disagreement on the Norwegian Supreme Court over a period of twenty years. This was a period in which disagreement rose, and in which the court decided to focus on cases with greater policy content. At the beginning of this period, the court was also appointed by the government. These are all factors which might reasonably have been expected to lead to the same left–right preferences which structure Norwegian politics expressing themselves on the court. Instead we have found that the single-best interpretation of division on the court lies in judges' differing approaches to the claims of public bodies, and the appropriate amount of deference to show to these authorities. These more-or-less deferential attitudes are structured by judges' personal attributes – former academics are much less deferential than career judges or lawyers from private practice. In this sense, 'who sits' matters. However, the system used to appoint seems not to matter, for we observed no significant differences before and after the 2002 reform to the appointment system.

The first implication of this research is that it is possible to explain judges' behaviour by drawing inferences about their preferences, and that explaining judicial behaviour using the techniques of item response theory does not mean assuming that judges are merely legislators in robes. We have not found that judges are divided between left- and right-wing judges, in the way that we might if judges were actually legislators in robes. The division that we have found – between judges who are (comparatively) deferential towards public authorities and judges who are not – is based on disagreement that is partly political and partly legal. Although our method of analysis has in the past often been used to substantiate claims about judges having strong political preferences, nothing in the method determines such a finding, and we expect that, as the comparative analysis of judicial disagreement in Western Europe develops, other courts will soon be shown to demonstrate similar, less obviously political patterns of disagreement.

Second, our findings provide support for a personal attributes model of judicial behaviour, with particular support for an account emphasizing judges' career trajectories. In this paper we are not able to identify the mechanism that makes academics less deferential, but past research has shown that former academics on international courts are more activist (Bruinsma 2006; cf. Jodoin 2010). Our finding therefore builds on past research.

Third, our findings have significant implications for institutional reform. The 2002 reform to appointments to the Supreme Court was intended to depoliticize appointments. We have shown that it did not matter, not least because there was no strong appointer–appointee relationship to begin with. We acknowledge that the 2002 reform might not have had an effect because Norwegian politicians did not need constraining, and therefore that the effects of moving from ministerial to merit-based appointment may vary depending on the initial conditions. Like Weiden (2011), we would suggest that selection culture, rather than formal appointment procedures, matters for the link between appointments and behaviour. This finding urges researchers to think twice before employing justices' appointing authorities as a proxy for their ideological preferences.

We see this article as part of a broader cumulative process of identifying and characterizing the bases of judicial disagreement on West European apex courts. This process is a necessary precursor to explaining why some courts are divided into left- and right-wing blocks whilst other courts are divided by matters that seem more exclusively legal in nature. Given the broader cultural similarities between the countries, we expect that disagreement in other Scandinavian apex courts can be characterized in the same way – but this would merely accentuate a puzzle we have already identified, namely why apex courts which at one point in time have been subject to fairly direct ministerial appointment should have resisted politicization in this way. Our findings thus call for further research into the conditions under which judicial appointment systems shape the bases of judicial disagreement on apex courts.

ACKNOWLEDGEMENTS

We thank Gunnar Grendstad, the SPS editor and anonymous reviewers for helpful feedback. Data and code to replicate our results is available at: https://doi.org/10.7910/DVN/DEVMTI.

REFERENCES

Alter, K. J. 2010. *Establishing the Supremacy of European Law: The Making of an International Rule of Law in Europe*. Oxford: Oxford University Press.

Arvind, T. T. & Stirton, L. 2016. 'Legal Ideology, Legal Doctrine and the UK's Top Judges', *Public Law*, *2016*(July), 418–36.

Bentsen, H. L. 2018. 'Court Leadership, Agenda Transformation, and Judicial Dissent: A European Case of a "Mysterious Demise of Consensual Norms"', *Journal of Law and Courts*, *6*(1), 189–213.

Bergara, M., Richman, B. & Spiller, P. T. 2003. 'Modeling Supreme Court Strategic Decision Making: The Congressional Constraint', *Legislative Studies Quarterly*, *28*(2), 247–80.

Bjørnebekk, O. & Johannesson, M. 2016. 'Domstolr', *GitHub Repository*. https://github.com/bearcreek/domstolr

Brace, P., Yates, J. & Boyea, B. D. 2012. 'Judges, Litigants, and the Design of Courts', *Law & Society Review*, *46*(3), 497–522. https://github.com/bearcreek/domstolr

Brouard, S. & Hönnige, C. 2017. 'Constitutional Courts as Veto Players: Lessons from the United States, France and Germany', *European Journal of Political Research*, *56*(3), 529–52.

Bruinsma, F. J. 2006. 'Judicial identities in the European Court of Human Rights', in Van Hoek, A., Hol, T., Jansen, O., Rijpkema, P. & Widdershoven, R. eds, *Multilevel Governance in Enforcement and Adjudication*. Intersentia, 203–40.

Bårdsen, A. 2017. 'The Norwegian Supreme Court as the Guardian of Constitutional Rights and Freedoms'. Presentation at "Norway in Europe", Centre for European Law, Oslo, 18 September. https://www.domstol.no/globalassets/upload/hret/artikler-og-foredrag/supreme-court---constitutional-rights---bardsen18092017.pdf

Bårdsen, A. 2018. 'Deciding What to Decide: The Filtering Mechanism in the Norwegian Supreme Court'. Presentation at Icelandic Icelandic Bar Association, Reykjavik, 16 February. https://www.domstol.no/globalassets/upload/hret/artikler-og-foredrag/deciding-what-to-decide---bardsen-23.02.2018.pdf

Carroll, R. & Tiede, L. 2011. 'Judicial Behavior on the Chilean Constitutional Tribunal', *Journal of Empirical Legal Studies*, *8*(4), 856–77.

Clinton, J., Jackman, S. & Rivers, D. 2004. 'The Statistical Analysis of Roll Call Data', *American Political Science Review*, *98*(2), 355–70.

Coroado, S., Garoupa, N. & Magalhães, P. C. 2017. 'Judicial Behavior under Austerity: An Empirical Analysis of Behavioral Changes in the Portuguese Constitutional Court, 2002–2016', *Journal of Law and Courts*, *5*(2), 289–311.

De Jaegere, J. 2019. *Judicial Review and Strategic Behaviour: An Empirical Case Law Analysis of the Belgian Constitutional Court*. Cambrigde: Intersentia.

Epstein, L. & Knight, J. 2013. 'Reconsidering Judicial Preferences', *Annual Review of Political Science*, *16*, 11–31.

Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y.-S. 2008. 'A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models', *The Annals of Applied Statistics*, *2*(4), 1360–83.

Grendstad, G., Shaffer, W. R., Skiple, J. K. & Waltenburg, E. N. 2015a. 'Statsvennlighet i norges høyesterett 1963–2013', *Norsk statsvitenskapelig tidsskrift*, *31*(4), 305–32.

Grendstad, G., Shaffer, W. R. & Waltenburg, E. N. 2015b. *Policy Making in an Independent Judiciary: The Norwegian Supreme Court*. Colchester: ECPR Press.

Hall, M. G. 1985. 'Docket Control as an Influence on Judicial Voting', *Justice System Journal*, *10*, 243–55.

Hanretty, C. 2012. 'Dissent in Iberia: The Ideal Points of Justices on the Spanish and Portuguese Constitutional Tribunals', *European Journal of Political Research*, *51*(5), 671–92.

Hanretty, C. 2014. 'The Bulgarian Constitutional Court as an Additional Legislative Chamber', *East European Politics and Societies*, *28*(3), 540–58.

Hanretty, C. 2015a. 'Judicial Disagreement Need Not Be Political: Dissent on the Estonian Supreme Court', *Europe-Asia Studies*, *67*(6), 970–88.

Hanretty, C. 2015b. 'The Appointment of Judges by Ministers: Political Preferment in England, 1880–2005', *Journal of Law and Courts*, *3*(2), 305–29.

Hirschl, R. 2009. *Towards Juristocracy: The Origins and Consequences of the New Constitutionalism*. Cambridge, MA: Harvard University Press.

Ho, D. E. & Quinn, K. M. 2010. 'How Not to Lie with Judicial Votes: Misconceptions, Measurement, and Models', *California Law Review*, *98*(3), 813–76.

Hönnige, C. (2009) The Electoral Connection: How the Pivotal Jugde Affects Oppositional Success at European Constitutional Courts. *West European Politics*, *32*(5), 963–984.

Jodoin, S. 2010. 'Understanding the Behavior of International Courts: An Examination of Decision-making at the ad hoc International Criminal Tribunals', *Journal of International Law and International Relations*, *6*, 1.

Keller, H. & Sweet, A. S. 2008. *A Europe of Rights: The Impact of the ECHR on National Legal Systems*. Oxford: Oxford University Press.

Kinander, M. 2016. 'Fra tilbakeholdenhet til aktivisme: Nyere utviklingslinjer i forholdet mellom rett og politikk i høyesterett', *Lov og Rett*, *55*(3), 141–64.

Kjønstad, A. 1997. 'Trygderettigheter, grunnloven og høyesterett', *Lov og Rett*, *36*(5), 243–92.

Kjønstad, A. 1999. 'Er høyesterett statsvennlig?', *Lov og Rett*, *38*(2), 97–122.

Lund, K. 1987. 'Kontroll av staten i statens egne domstoler', *Lov og Rett*, *26*(4), 211–27.

Magnussen, A.-M. 2005. 'The Norwegian Supreme Court and Equitable Considerations: Problematic Aspects of Legal Reasoning', *Scandinavian Political Studies*, *28*(1), 69–89.

Malecki, M. 2012. 'Do ECJ Judges All Speak with the Same Voice? Evidence of Divergent Preferences from the Judgments of Chambers', *Journal of European Public Policy*, *19*(1), 59–75.

Martin, A. D. & Quinn, K. M. 2002. 'Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953–1999', *Political Analysis*, *10*(2), 134–53.
Martinsen, D. S. 2011. 'Judicial Policy-making and Europeanization: The Proportionality of National Control and Administrative Discretion', *Journal of European Public Policy*, *18*(7), 944–61.
Melton, J. & Ginsburg, T. 2014. 'Does De Jure Judicial Independence Really Matter? A Reevaluation of Explanations for Judicial Independence', *Journal of Law and Courts*, *2*(2), 187–217.
Mestad, O. 2015. 'Dissensens alvor: Om nyare dissensar i høgsterett', in Schei, T., Skoghøy, J. E. A. & Øie, T. eds, *Lov, sannhet, rett: Norges høyesterett 200 ar*. Oslo: Universitetsforlaget, 222–253.
Pellegrina, L. D., De Mot, J., Faure, M. & Garoupa, N. 2017. 'Litigating Federalism: An Empirical Analysis of Decisions of the Belgian Constitutional Court', *European Constitutional Law Review*, *13*(2), 305–46.
Pellegrina, L. D. & Garoupa, N. 2013. 'Choosing between the Government and the Regions: An Empirical Analysis of the Italian Constitutional Court Decisions', *European Journal of Political Research*, *52*(4), 558–80.
Popelier, P. & Bielen, S. 2019. 'How Courts Decide Federalism Disputes: Legal Merit, Attitudinal Effects, and Strategic Considerations in the Jurisprudence of the Belgian Constitutional Court', *The Journal of Federalism*, *49*(4), 587–616.
Robberstad, A. 2016. 'Viljen til å skape lov og grunnlov', *Lov og Rett*, *55*(1), 49–64.
Segal, J. A. & Spaeth, H. J. 2002. *The Supreme Court and the Attitudinal Model Revisited*. Cambridge: Cambridge University Press.
Shaffer, W. R., Grendstad, G. & Waltenburg, E. N. 2015. 'Is Diversity Just For Show? Diversity and Appointment to the Norwegian Supreme Court 1945–2011', *Raetferd*, *38*(1), 18–42.
Shapiro, M. & Stone, A. 1994. 'The New Constitutional Politics of Europe', *Comparative Political Studies*, *26*(4), 397–420.
Shapiro, M. & Sweet, A. S. 2002. *On Law, Politics, and Judicialization*. Oxford: Oxford University Press.
Skiple, J. K. 2015. 'Ei konkret vurdering i kvart enkelt tilfelle. Ein fleirnivåanalyse av økonomisk stemmegiving i høgsterett i tidsrommet 1991–2011', *Norsk Statsvitenskapelig Tidsskrift*, *31*(4), 278–304.
Skiple, J. K., Grendstad, G., Shaffer, W. R. & Waltenburg, E. N. 2016. 'Supreme Court Justices' Economic Behaviour: A Multilevel Model Analysis', *Scandinavian Political Studies*, *39*(1), 73–94.
Stan Development Team. 2018. 'RStan: The R interface to Stan'. Available online at: http://mc-stan.org/
Sunde, J. 2015. *Høgsteretts historie 1965–2015. At dømme i sidste instans*. Bergen: Fagbokforlaget.
Tate, C. N. 1981. 'Personal Attribute Models of the Voting Behavior of Us Supreme Court Justices: Liberalism in Civil Liberties and Economics Decisions, 1946–1978', *American Political Science Review*, *75*(2), 355–67.
Valdini, M. E. & Shortell, C. 2016. 'Women's Representation in the Highest Court: A Comparative Analysis of the Appointment of Female Justices', *Political Research Quarterly*, *69*(4), 865–76.
Voeten, E. 2007. 'The Politics of International Judicial Appointments: Evidence from the European Court of Human Rights', *International Organization*, *61*(4), 669–701.
von Eyben, W. 1956. 'Judicial Law making in Scandinavia', *The American Journal of Comparative Law*, *5*(1), 112–15.
Weiden, D. L. 2011. 'Judicial Politicization, Ideology, and Activism at the High Courts of the United States, Canada, and Australia', *Political Research Quarterly*, *64*(2), 335–47.
Weinshall, K., Sommer, U. & Ritov, Y. 2018. 'Ideological Influences on Governance and Regulation: The Comparative Case of Supreme Courts', *Regulation & Governance*, *12*(3), 334–52.
Wiklund, O. 2008. 'The Reception Process in Sweden and Norway', in Kellen, H. & Stone Sweet, A., eds, *A Europe of Rights: The Impact of the ECHR on National Legal Systems*. Oxford: Oxford University Press, 165–228.
Windett, J. H., Harden, J. J. & Hall, M. E. K. 2015. 'Estimating Dynamic Ideal Points for State Supreme Courts', *Political Analysis*, *23*(3), 461–69.

APPENDIX

Table A1. Ordinary least squares regression models of judicial ideal points as a function of personal attributes and career trajectory. Model 4 uses the primary career of each judge relative to the baseline category (legal academic); Models 5 and 6 use non-exclusive codings of whether each judge has ever worked in a particular role

	Model 4	Model 5	Model 6
(Intercept)	0.508	0.219	0.205
	(0.381)	(0.437)	(0.388)
Born in Oslo	0.124	–0.032	
	(0.283)	(0.297)	
Career: Attorney (vs. legal academic)	–1.064*		
	(0.471)		
Career: Government adm.	–0.415		
	(0.423)		
Career: Judge	–1.033		
	(0.548)		
Interim judge	0.152	0.088	
	(0.647)	(0.657)	
Female	–0.586	–0.375	
	(0.322)	(0.308)	
Ever an attorney		–0.757**	–0.680*
		(0.279)	(0.289)
Ever a judge		–0.196	–0.285
		(0.314)	(0.273)
Ever in government administration		0.072	
		(0.348)	
Ever legal academic		0.615	0.619
		(0.404)	(0.358)
Ever a government attorney			–0.235
			(0.303)
Ever in the government legislation dept.			0.069
			(0.296)
Ever a public prosecutor			–0.098
			(0.565)
Num. Obs.	47	46	46
R^2	0.216	0.284	0.265
Adj. R^2	0.098	0.152	0.151
AIC	132.4	128.1	127.3
BIC	147.2	144.6	142.0
Log. Lik.	–58.217	–55.051	–55.666

*p < 0.05; **p < 0.01.

Table A2. Column (1) reports the results of the regression as estimated on all 47 judges. Column (2) reports the results of the regression as estimated on 28 judges who were either appointed and served before the 2002 appointments reform, or who were appointed after the reform

	(1)	(2)
(Intercept)	–0.172	–0.031
	(0.248)	(0.456)
Appoint govt. SD	0.058	0.070
	(0.296)	(0.465)
Post-2002 reform	–0.042	–0.191
	(0.290)	(0.465)
N	47	28
R^2	0.001	0.007
Log Lik	–63.312	–41.539
AIC	134.623	91.079

*p < 0.05; **p < 0.01; ***p < 0.001.