Gender in the Supreme Court of Norway

Judicial behaviour in child custody cases

1968 – 2011

Master thesis

Kurt-Rune Bergset

Spring 2013

UNIVERSITY OF BERGEN

Department of comparative politics

Abstract

This thesis explores how gender affects the votes of Norwegian Supreme Court judges in cases of disputed child custody in the period from 1968 to 2011.

A typology of four common approaches to the study of gendered behaviour is tested. The Different Voice approach assumes that men and women reason in a fundamentally different way. Within the Representative approach, gender groups are perceived as social classes. The Informational approach assumes that the differences between male and female judges are due to information they gain from social interaction throughout life. Finally, the Organisational approach assumes that any initial gender differences are subdued by professional experience.

Hypotheses inferred from each theoretical approach are tested by employing multi-level logistic regression analysis. The dependent variable is the individual judge's propensity to vote for the feminine party. The data is organised in a two-level hierarchical model, where factors related to judges and their votes are placed at level 1. Factors related to panel composition or case particularities are placed at level 2. Because this structure does not control for inter-individual dependencies, an important part of the analysis is to identify and control for the effect of critical actors, i.e. judges who vote consistently either in favour or disfavour of the female party. An important methodological conclusion is that this seems to be a prudent manner of controlling for inter-individual dependencies and expose spurious effects.

The data consists of 750 votes in 150 child custody cases. This is the entire population of cases between 1968 – the year justice Lilly Bølviken was appointed the first female judge in the Norwegian Supreme Court – and 2011.

The analysis shows that female judges are less likely to vote in favour of the feminine party. Gender only explains a fraction of the differences in judges' voting pattern. It is, however, likely to affect judges in their exercise of judicial discretion and in cases of dissent. Both the Informational and Organisational approaches seem ill fit to explain the observed gendered behaviour in these cases. Although both have some weaknesses, both the Different Voice and Representational approaches seem fruitful in this respect.

Acknowledgements

Writing this thesis has been one of the most challenging tasks I have taken upon myself. The learning curve has been steep both with regard the legal method, gender theory and quantitative analysis. I could not have done it without assistance.

First of all, I am very grateful to my supervisor, professor Gunnar Grendstad, without whom I would never have been writing a thesis about this subject. His involvement of students in contemporary research has been a great inspiration to more people than me. As a supervisor, Grendstad has been competent, helpful and flexible far beyond expectation. In addition, the opportunity I had to be employed as research assistant for the project *Judicial Behaviour in the Supreme Court of Norway* has given me invaluable experience and insight. For this, I thank Grendstad and his colleagues, professors Bill Shaffer and Eric Waltenburg.

Special thanks go to former Supreme Court justice Kirsti Coward for valuable and interesting conversations during a study trip to Oslo in 2012. Also deserving thanks is professor of legal history, Jørn Øyrehagen Sunde for insights and open-minded participation in our seminars.

From Marius Svendsen and head librarian Jørn Wangensten Ruud I received invaluable help in developing relevant keywords to extract data from Lovdata. I am also thankful to the Centre for Women's and Gender Research at the University of Bergen for letting me present my thesis at the Masterstafetten conference, and to the Institute of Comparative Politics for letting me present my thesis during a research meeting, and to both for valuable feedback.

To all fellow students who have made my life a good place the last year, my colleagues at Ullaloftet and my colleagues in the Doranoh research group: I am grateful for all assistance, intelligent distractions and joy! There are too many of your names to fit this page!

Finally, huge thanks to Elin, who has been a moral support, academic guide, good company, proof reader, etc. All in one, competent and without any utterance of complaint.

Thank you all!

On translations, abbreviations and the likes

For translations of names of institutions and organisations to English, the official translations from these organisations' web pages have been used, where available. These sources are not referred, however. For legal concepts, Åge Lind's English terms list in *Jusleksikon* by Jon Gisle et al. (2010) has been an important source of translations, and also not consistently referred. Where concepts have already been established in the relevant literature, the translations will diverge from Lind's terms. An example of this is the translation of the concept *reelle omsyn*, which Lind translates to *public policy considerations*. This is translated to *equitable considerations* by, among others, Grendstad et al. (2011b). To promote conceptual clarity in such cases, the priority falls on the translation used in relevant literature.

The Norwegian terms in the thesis are written in the *Nynorsk* variant of Norwegian, in which e.g. the Supreme Court is called (*Noregs) Høgsterett*, while it would be called (*Norges) Høyesterett* in the *Bokmål* variant.

List of tables and figures

Gender in the Supreme Court of Norway

1 Introduction

In a commentary in Aftenposten, one of the prime Norwegian newspapers, journalist Inge D. Hanssen (2010) summarised the process of a of disputed child custody case after the split between a girl's two parents[1]. The case was one out of very few such cases to have gone the long way through three court levels, from the district court, *tingretten*, via the court of appeals *lagmannsretten*, to the Supreme Court of Norway, *Noregs Høgsterett*.

An important property of these cases, as Hanssen describe them, is the high level of judicial discretion needed to cast a verdict[2]. The judges are bound to evaluate what is in "the best interest of the child", a concept criticised for its ambiguity (e.g. Skivenes, 2010). The few concrete factors upon which the judges can evaluate the goodness of an outcome are normally distorted by the high level of conflict and accusations common to child custody cases.

Despite this, the judges of the court are bound to reach a conclusion, and in matters of judicial discretion, they are also theoretically bound by the legal method to seek a conclusion they find morally good (see Eckhoff and Helgesen, 1997:357-60). Journalist Hanssen's observation was that in the particular case at hand, the judicial discretion had seemingly been exercised differently at all three levels of court. Consequently, the outcome of the case had changed completely at every appeal. What more, Hanssen thought he saw a simple pattern in the exercise of discretion:

> *"Could it be that judges' gender matters when discretion is exercised? The district court judge is female. She found for mother. The three appellate judges who found that the daughter ought to live with her father are all male. The five Supreme Court judges, who finally found for mother, are all female. Coincidental?"*
> *(Hanssen, 2010, my translation)*

The present thesis provides an answer to Hanssen's open question. An analysis of the entire population of child custody cases before *Høgsterett* between 1968 and 2011 shows that the pattern Hanssen has seen is indeed coincidental. The tendency in the Supreme Court of Norway is, in fact, opposite of what one would have expected from this particular case.

[1] The case in question has the reference number Rt-2010-216 and is one of 150 cases included in this analysis.
[2] The high need for judicial discretion is likely to be a property of most Supreme Court cases.

1.1 The research question and its contribution

How does gender affect the voting pattern of Norwegian Supreme Court judges in cases of disputed child custody?

Since 2010, research on judicial behaviour in Norway has been expanded to include advanced statistical analyses of justices' voting pattern in the Supreme Court of Norway, *Høgsterett*. Previous statistical analyses have mainly focused either on dissent or on ideological voting in cases where *the public interest* is a party. In most of these analyses, gender is included either in the form of individual judges' gender or as a measure of the gender division in each presiding panel.

Where Grendstad et al. (2011b) found no significant effect of individual judges' gender in their analysis of non-unanimous cases in *Høgsterett,* more recent studies that include larger selections of cases have shown that gendered behaviour in the Supreme Court of Norway is an issue that deserves further analysis. Jacobsen (2012) found that individual judges' gender might affect judges' propensity to vote in favour of the state in civil cases. Similarly, a study by Skiple (2012) indicates that the likelihood for a judge to vote in favour of the public interest is higher in panels where the majority of judges are female.

This thesis makes three major contributions to the field of judicial behaviour in Norway. First, the thesis contributes with a theoretical expansion of the field of judicial behaviour, by introducing and testing four general theories of gendered behaviour in the court. Second, the thesis contributes with a wider empirical understanding of how gender affects judges' votes in the Supreme Court of Norway. Finally, the thesis contributes to the methodology of the field, by introducing a mechanism to test for the impact of critical actors in judicial panels.

The judicial branch in Norway has the prerogative of legal interpretation. Høgsterett is the court of last resort, and knowledge of the extent to which non-legal factors might affect the voting pattern of the judges in this court is therefore important. This analysis sheds some light on one of these factors.

1.2 Disposition

Chapter 2 provides a theoretical base for the thesis. The chapter starts with a presentation of a fourfold typology of the most common approaches to gendered judicial behaviour. Based on this typology, hypotheses are inferred from each general approach, to allow for empirical testing.

Chapter 3 explains the methodological rationale behind the thesis' empirical analysis. The first part of the chapter, section 3.1, is a short review of the methodological development of the field, with the purpose of positioning the thesis in the literature. Following this, in section 3.2, is an extensive presentation of logistic multi-level regression analysis. Here, central technical concepts related to the analysis are introduced and elaborated upon. Finally, in section 3.3, a four-step analytical strategy is presented. In this final section, the course of the analysis is explained step-by-step.

An overview of the data the analysis builds upon can be found in *Chapter 4*. The first part of this chapter, section 4.1, introduces the main data sources of the analysis. The following section 4.2 provides a detailed description of the three phases of data collection conducted in preparation of the analysis. In the third part of the chapter, section 4.3, potential sources of bias are addressed, before the analysis' variables are given operational definitions in section 4.4.

The analysis itself is described in *Chapter 5*, providing an empirical evaluation of the hypotheses derived from the typology of gendered behaviour. The first part of the chapter, section 5.1 is a descriptive analysis of each variable in the analysis. In the following, section 5.2, the analysis is conducted as described in the analytical strategy presented in chapter 3.

The conclusions and implications of this empirical evaluation are discussed in *chapter 6*, along with suggestions for future research.

2 Theories of gendered judicial behaviour

The study of how gender affects the behaviour of judges is a highly specialised sub-genre of the vast field of judicial behaviour, in which the methodology and theories of the social sciences are implemented to study the behaviour of the upholders of the law.

Boyd et al. (2010) has developed a typology identifying four mutually exclusive conceptions of the causes and effects of gendered behaviour within this literature. These are the Different Voice approach; the Representative approach; the Informational approach, and finally; the Organisational approach. This typology makes out the basic theoretical structure of the present thesis, and in the following, the typology will be presented and adapted to the context of the Norwegian judicial system.

Table 2.1 on the following page provides a summary of the theoretical basis for the present thesis. The four general approaches presented in that table are explained and elaborated upon in section 2.1.

In addition to the four general approaches, three additional concepts, *Critical actors*, *critical mass* and *tokens* are also presented and fitted into the typology in Table 2.1. These three concepts are elaborated upon in section 2.2.

Finally, expected effects in the form of testable hypotheses derived from the four theoretical approaches are presented in section 2.3.

Table 2.1: Typology of four approaches to gendered behaviour.

Case fields:	*All issues*	*Gender-related issues*		*No issues*
Theoretical approaches	**Different voice** *(Gilligan, 1982)*	**Representational** *(Pitkin, 1967)*	**Informational** *(Peresie, 2005, Boyd et al., 2010)*	**Organisational** *(Steffensmeier and Herbert, 1999)*
Summarized theories	The male and female perspective is *fundamentally different* in *all types of issues*.	Gender groups are perceived as *social classes*. In *gender-related issues,* judges will favour their own social group's interest.	Through socialisation, men and women gain *different* information about *gender-related issues* and will thus perceive these differently.	Men and women are taught the same curriculum and have identical jobs. Thus, gender should not affect their professional behaviour.
LEVEL 1: Individual effects of gender	Women *will* vote significantly different from men in all case fields.	Women *will* vote significantly different from men within *gender-related issues*.	Women *might* vote significantly different from men within *gender-related issues*.	*None expected*.
LEVEL 2: Collegial effects of gender	*None expected*. Men and women approach law completely differently.	*None expected.* A logical consequence is, however, that some (but not a majority of) men also *might* represent "the female side" within gendered issues.	*Yes*. Through deliberation, men and women *will* influence each other's perspectives and behaviour.	*None expected.*
Critical actors	*None expected* However, they are not logically excluded insofar as they have a voting pattern similar to that of their own gender group's norm.	*Yes*. We *might* expect some persons to more clearly represent their class or group. This *might* also include individuals from the opposite gender. We should not expect critical actors to have any no collegial effects, however.	*Yes.* Any actor entering a deliberative forum can bring an informational background along, which affects their perception. Notably, this information can be shared through deliberation, and we should therefore expect collegial effects.	*None expected. However, critical actors are not logically excluded.*
Critical mass	*None expected*. However, tokenism has been used to explain lack of female divergent behaviour in some literature in this tradition.	*Not excluded.* Assuming rationality, we can expect any gender effect to be strengthened where women have reached a *critical mass* where they are able to win through.	*Not excluded*. Assuming rationality, we can expect any gender effect to be strengthened where women have reached a *critical mass* where they are able to win through.	*Not applicable*.
Supported models of judicial behaviour	*Attitudinal model* *See Segal and Spaeth (2002)*		*Strategic-internal model* *See Dyevre (2010)*	*The legal model* *See Gillman (2001) or Segal and Spaeth (2002)*

Main sources: *The main structure is provided by Boyd et al. (2010). Also Steffensmeier and Herbert (1999), Palmer (2001) and Allen and Wall (1993).*

2.1 Four approaches to gendered behaviour

2.1.1 The Different Voice approach

Inspired by Carol Gilligan's seminal book (1982), the Different Voice tradition of studying gendered judicial behaviour has its main purpose in testing the assumption that the male and female jurisprudences are different at their core. Gilligan, a psychologist, made the case that from the earliest time of childhood, men and women develop distinct world-views, which continue through adult life. In 1986, Sherry developed a theoretical framework for studying gendered judicial behaviour, drawing heavily on Gilligan's theories. She contends that the difference in male and female jurisprudence can be equalled to two paradigms, where "the masculine vision parallels pluralist liberal theory [while] the feminine vision is more closely aligned with classical republican theory" (1986:543). Where women seek *connection*, *contextuality* and *responsibility*, men seek *autonomy*, *abstraction* and *rights*. A feminine jurisprudence is thus assumed to be based on *communitarianism* rather than the *individualism* of masculine jurisprudence. In Behuniak-Long's (1992:427) words, a feminine jurisprudence rejects "an adversarial, dichotomous, zero-sum game perspective of the issues [...M]aternal legal thinking can suggest innovative resolutions that offer concessions to both sides." (see also Palmer, 2001:94). This feminine perspective is assumed to "extend well beyond areas traditionally seen as affecting women, and in fact encompass all legal issues" (Sherry, 1986:581).

In Boyd et. al's (2010) typology of gendered judicial behaviour, the *different voice approach* thus assumes that *on the individual level*, we find significantly different behaviour between female and male justices across most (or all) areas of the law. On the *collegial level*, however, we should expect no significant differences, since male and female justices are unlikely to influence one another.

2.1.2 The Representational approach

The representational approach, a social classes theory, can be traced back to Pitkin's (1967) *"The Concept of Representation"*. The basic theoretical assumption is that female judges are representatives of women *as a social class*, and will work towards the common goals of their class – i.e. to further the goals of women's liberation or in other ways front the common interests of women (Boyd et al., 2010:390-1). Consequently, like in the *different voice* approach, we should assume that male and female judges behave differently. But it is only in

those areas-of-law where we can make the case that *women's interests* confronts either *men's* or *the established society's* interest that we can assume that female judges will rule in a manner different from that of male judges[3]. In this thesis, these kinds of issues will be conceptualised as *gender-related issues, gender-related cases,* or the like. Some examples Boyd et al. (2010:391) draws from the US context are abortion, affirmative action, sex discrimination in employment, and sexual harassment in the work place.

Assuming that women represent their own *class,* we must also assume that men represent theirs, and the purist interpretation of this model is similar to the different voice approach in that we expect no panel effects (men will not be affected by women's positions and vice versa). A necessary logical consequence of this approach is, however, that also some men can function as representatives of women *as a class* (and vice versa). This can be documented through history, where e.g. British liberalists like Jeremy Bentham and John Stuart Mill promoted of the movement for women's suffrage, where dramatists like Henrik Ibsen promoted women's liberation through seminal plays like *A Doll's House*. Notably, in 1884, social-liberalist (and man) Hagbard Emanual Berner was elected the first ever chairperson of the newly founded *Norwegian Association for Women's Rights*[4], promoting women's rights.

2.1.3 The Informational approach

The basic assumption in this approach, as it is presented by (Boyd et al., 2010:391) is that women – as the "newcomers" on the bench – possess "unique and valuable" *information* that men do not have. As with the representational approach, we thus assume the individual gender effect to be limited to a certain set of *gender-related* areas of law where women and men might possess different information. When modelling and empirically testing the representational and the informational approaches against each other, the main difference between the two is that with an informational approach we have to assume so-called *panel effects* or *collegial effects,* that female judges, bringing their gender-specific *information* to the bench will affect male judges' perspectives[5] (and vice versa).

[3] Also, consequently, however more or less untouched by literature, we should assume that men act as representatives of their own class. Their class is in this perspective assumed to be the dominant class, and thus also the representative of *established society*.

[4] *Norsk kvindesaksforening*.

[5] These kinds of *panel effects* have only recently become the focus of attention in analyses of judicial behaviour, but they have been found in several analyses, see i.e. Peresie (2005), Baldez et al. (2006), Cameron and Cummings (2003), Ostberg and Wetstein (2007), Sullivan (2002). In the Norwegian context no-one has analysed this in a gender perspective, but for reference see also Jacobsen (2012), Skiple (2012) and Bentsen (2012)

Boyd et al. (2010:391, note 7) finds it reasonable to also limit the number of *gender-related cases* to one very specific area of law, *gender discrimination in the employment context*. They specify, however, that this very strict interpretation has been a debated issue when presenting their paper to professional audiences. Arguing their case, they point to public opinion data, which indicates no significant differences between male and female perception of e.g. abortion, but "considerable differences" with respect to gender discrimination in the work force. However, because the Informational approach is the only approach that can be identified by collegial effects of gender, Boyd et al. leaves the debate without any further arguments.

This leaves an unanswered question as to why *only* gender discrimination in the work place should be a gender-related issue and not e.g. violence or child-care, which are areas where – in general – men and women seem to have both differing opinions and roles. Hirsch (2010:22-23) shows that there are significant differences between men and women as to whether they commit violence and which kinds of violence they commit. Equally, there are significant differences as to whether they fear violence and what sort of violence they fear. Similarly, when it comes to child-care, Vaage (2012) shows that, although differences between the sexes are evening out, the general trend is still that women and men have clearly different roles with regard to the household and child-care[6].

Summing up, Boyd et al.'s argument for limiting the number of potential gender related cases to only one case field can also be used for other case fields. As such, in the present thesis, the concept of gender-related issues is understood to be equally wide in both the Informational approach and in the Representational approach, and the main difference between the two is found in whether we can observe collegial effects of gender.

2.1.4 The Organisational approach

If the Different Voice approach could be classified as a typical maximalist approach to gendered behaviour, the Organisational approach should by understood as a typical minimalist approach (Steffensmeier and Herbert, 1999). On one side, any general differences

[6] In 2010, where women spend 28 per cent more time working in the household than men, men spent 38 per cent more time on income work than did women (Vaage, 2012:13-14). Similar differences are found for the mean time men and women spend with their children (Vaage, 2012:219, 224).

found between men and women are assumed to simply reflect the influence of the constraints and opportunities to which a person has been exposed. On the other side, in a professional setting, like a judicial collegium, we should assume that "professional training and identical constraints imposed by and rules overcome any biological, psychological, or experience-based differences between the sexes" (Steffensmeier and Herbert, 1999:1165). As long as female and male judges are taught the same curriculum in law school and have similar jobs received through the same channels, gender should not have any effect on how they exercise their profession.

2.2 Three additional concepts: Tokens, critical mass and critical actors

Suggestions have been made that women assimilate male attitudes over time[7]. "Women who become federal judges may simply adapt to the requirements of the male oriented power structure [...] They may not find an opportunity to depart from those rules to express a different voice" (Davis (1994:171), see also Menkel-Meadow (1989:313-14)). This entails that the first women on the bench are so-called tokens (see Martin and Pyle (2000:1214), footnote 62, 68-70). Justice Sandra Day O'Connor, the first female judge on the US Supreme Court, justice Bertha Wilson on the Canadian Supreme Court and justice Lilly Bølviken in Høgsterett in Norway are assumed to have adapted to a male set of thinking and behaving to achieve their positions. The tokenist perspective thus entails that we can expect little traceable gender effect in the behaviour of these three judges who are likely to have adapted to the male paradigm[8] of jurisprudence.

One logical consequence of the token argument is the critical mass argument: As the sheer number of female judges rise and reach a critical mass, "assimilation may be replaced by transformation" (Davis, 1994:172) and female judges will be able and willing to act on their female jurisprudence. According to Martin and Pyle (2000:1219), "studies indicate that as women's numbers move beyond tokenism, and as younger females who are educated after the women's movement become judges, differences based on gender emerge more clearly." Equally, McCall and McCall (2007) finds that "women justices, controlling for institutional, political, and legal constraints, are more likely to rule in favour of the criminal defendant than their male brethren in cases decided after 1991 but not before." They also "conclude the

[7] Note that Ostberg and Wetstein (2007) found no evidence for attitudinal assimilation *during the judge's time on the bench* in the Supreme Court of Canada, but in the tokenist perspective, this assimilation has probably happened a long time before the women reach the highest and most prestigious court in the country.
[8] Sherry (1986) introduced the concept of a male paradigm of jurisprudence.

influence of gender may be evident in a wider variety of cases than those dealing with women's lives but that this influence is dependent upon the existence of a critical mass of women on the state court benches."

Another logical consequence of the token argument is the critical actors argument. If the early judges are likely to have made their position through adapting to a male paradigm, we could equally assume that some judges (perhaps in a second wave of female judges) would be strongly oriented towards a "female jurisprudence". Promoting a *critical actor*, such as the highly outspoken and notorious dissenter, justice Claire L'Heureux-Dubé who served on the Supreme Court of Canada from 1987 to 2002, should thus introduce a voice critical of the male paradigm of the court.

In the perspective provided to us by the Different Voice approach, critical actors are problematic. With this approach, we assume that male and female reasoning is fundamentally different. As such, we should assume that *if* critical actors exist, these judges would exhibit a voting pattern similar to the general pattern in their own gender group. Specifically, if male judges are generally found to vote more in favour of the non-female party than female judges do, we expect a male critical actor to show the same individual pattern.

The concept of critical actors is a less problematic concept within the Representational approach, where differences in male and female jurisprudence are assumed to be based in class orientation. Certain actors are likely to represent a class or group more clearly, and, as has been discussed in section 2.1.2, this can include actors who belong to a different class or group. As such, we can expect critical actors, and we can expect both male and female judges to stand out as critical actors for a "female jurisprudence". Because the male paradigm should be assumed the norm of the court, most critical actors (who by definition diverge from the norm) should be assumed to stand out as critical actors *for* the female party.

Finally, within the Informational approach we should expect critical actors to behave on the basis of their informational background. Within this approach, we assume that any gender effect we find is related to the informational background common to the members of the two gender groups. As such, any one of the judges who bring with them particular information on a subject will potentially act upon this information. Additionally, insofar as a judicial panel is a deliberative forum, we should expect critical actors to affect their peers voting pattern by

sharing their particular perspectives. Thus, this is the only approach that opens for general collegial effects of critical actors.

The introduction of the concept of critical actors also has a methodological function, as a control mechanism for *inter-individual dependencies*. In that respect, the subject will be further elaborated upon in chapter 3, section 3.2.4.

2.3 Expected effects – hypotheses

Table 2.2: Summary of the expected effects within four approaches to gendered behaviour

Theory	Measure/variable	Expected effect
Expected individual effects		
Different voice, Informational, Representational *Disfavouring:* Organisational	**Gender**	*H1: "Male and female judges vote significantly different from each other"*
Representational *Disfavouring:* Organisational	**Gender**	*H2: "Female judges are significantly more likely than male judges to vote in favour of the female party"*
Organisational *Disfavouring:* Different voice, Representational, Informational.	**Gender**	*H3: "Any difference found in the voting pattern of male and female judges is purely random"*
Informational, Representational *Disfavouring:* Organisational	**Critical actors**	*H4: "Some individual judges – so-called critical actors – will vote consistently either in favour or disfavour of the female party throughout all cases"*
Representational *Disfavouring:* Organisational	**Critical actors**	*H5: "A large majority of the judges who are identified as critical actors will vote consistently in favour of the female party throughout the cases"*
Different voice *Disfavouring:*	**Critical actors**	*H6: "Any judges identified as critical actors will vote consistently in favour of their gender group's preferred party"*
Expected collegial/panel effects		
Informational, Representational, *Disfavouring:* Different Voice	**Female majority**	*H7: "Any individual gender effect will be strengthened in panels with a critical mass – a majority – of women."*

Informational *Disfavouring:* Representational, Different voice, Organisational	**Gender percentage**[9]	*H8: "Controlling for any individual effect of gender, the voting pattern of judges is affected by the gender balance of the panel in which they vote."*
Informational *Disfavouring:* Representational, Different voice, Organisational	**One woman present**	*H9: "Having at least one woman present will affect the voting behaviour of all judges in the panel."*
Informational *Disfavouring:* Representational, Different voice, Organisational	**Presiding judge female**	*H10: "Controlling for any individual effect of gender, the gender of the presiding judge will affect the voting pattern of other judges in the panel."*
Organisational, Representational Different voice *Disfavouring:* Informational.	**One woman present** **Female majority** **Gender percentage**	*H11: "The voting pattern of judges is not affected by the gender balance in the panel"*
Informational *Disfavouring:* Representational, Organisational, Different voice	**Critical actors present** **&** **Gender**	*H12: "Parts of any individual gendered voting pattern found in the analysis can be explained by the presence of a critical actor in the panel in which they cast their vote."*
Informational *Disfavouring:* Different voice, representational, organisational	**Critical actors present**	*H13: "The presence of a critical actor in a panel significantly raises the likelihood of a judge voting in favour of the critical actor's preferred party."*
Expected effects of case particularities		
Different voice	**Gender of opposing party's lawyer**	*H14: "Any individual gender effect is strengthened if the party which female judges are more likely to vote in favour of is represented by a female lawyer"*
Representational	**Opposing party is public/male**	*H15: "Any individual gender effect will be strengthened when controlling for whether the female party's opposing party is a male individual or a representative of the public."*
Informational *Disfavouring:* Representational, Different voice	**Time variables**	*H16: "Any individual gender effect will be alleviated by the number of years women has served alongside men in the court"*

[9] In this selection only one single panel consists of only women, so a variable accounting for all-women panels is excluded.

2.3.1 The individual level

Effects of gender on the individual level

The Different Voice approach proclaims that men and women have fundamentally different world-views and will reason in a different manner in all case fields. However, the approach does not provide any clear indication as to which general direction we might expect e.g. female judges to vote in cases of child custody.

Insofar as the selection of cases is limited to gendered issues, the Informational approach also suggests that female and male judges might reach differing conclusions in the same cases. Here, however, this is assumed to be because of their informational background and not on fundamentally differing world-views. Similar to the Different Voice approach, the Informational approach does not indicate any clear direction of the effect:

H1: "Male and female judges vote significantly different from each other"

The Representational approach, also limited to gendered issues, proposes that female judges will support the interests of women as a social group or a class. However, it differs in that it specifies a direction of the gender effect, insofar as we can assume that a representative of women's interests would vote in favour of the female party in cases where women's interest are at play.

H2: "Female judges are significantly more likely than male judges to vote in favour of the feminine party"

The Organisational approach differs from the previous three approaches in not expecting any significant differences between the voting patterns of male and female judges. This approach assumes that, insofar as the judges have been taught the same curriculum and have similar qualification, there should be no difference between the professional behaviour of male and female judges. Therefore, what we infer from the Organisational approach is in effect the null-hypothesis of the first two hypotheses.

H3: "Any difference found in the voting pattern of male and female judges is purely random"

Critical actors on the individual level

The concept of critical actors does not fit perfectly within the typology of four approaches to gendered judicial behaviour. The concept is in use within all the three traditions that expect gender effects, but only fits well logically within the Representational and Informational approaches.

Insofar as the Representational approach assumes, in a class perspective, that female justices will function as representatives of their gender's general interests, it is reasonable to assume that some judges are more class-conscious than others are. What ought to be noted is that the representational approach does not exclude the possibility that male judges function as class representatives for "the female class"[10].

Likewise, within the informational approach we must assume that individuals might bring particular information, such as particular life experiences, or the likes, to the bench. With this approach, individual critical actors can act on far wider grounds than just their gender, and both male and female judges can have individual experiences, or the likes, that affect their general attitudes to questions of child custody.

We can therefore infer the following common hypothesis for these two approaches.

H4: "Some individual judges – so-called critical actors – will vote consistently either in favour or disfavour of the female party throughout all cases"

Because the Representational approach assumes that judges are affected by their ideological support for certain social groups or classes, it is reasonable to assume that some actors are more engaged in this ideology, or more "class-conscious" than others are. As has been argued for earlier, acting as a representative of the "female class" would not necessarily be bound merely to female judges. Judges of both genders can be inclined to support a particular group they identify with or perceive to be in need of their support. Because the norm of the court should be the "male paradigm"[11], we should expect *most* critical actors – insofar as being a critical actor means having a voting pattern that disaccords with the normal pattern - to act as

[10] For a discussion, see section 2.1.2.
[11] See section 2.1.2.

representatives of the female interest. Based on this reasoning we can infer the following additional hypothesis on behalf of the Representative approach.

H5: "A large majority of the judges who are identified as critical actors will vote consistently in favour of the female party throughout the cases"

Because the Different Voice approach assumes that men and women have a fundamentally different way of reasoning, there is no reason to assume that we will find any particular critical actors in addition to the expected gender effect. However, the approach does not logically exclude the existence of critical actors, insofar as they vote in accordance with the trend predicted for their gender group.

H6: "Any judges identified as critical actors will vote consistently in favour of their gender group's preferred party"

Finally, on behalf of the Organisational approach, we should expect to find no critical actors diverging consistently from the norm. Like the Different Voice approach, however, this approach neither excludes the concept of critical actors. Insofar as judges are professional actors, any significant and consistent individual divergence should be understood in terms of professional disagreement and not in a gendered or social context. While inferring a hypothesis regarding critical actors on behalf of this approach is problematic, a rejection of the previous three hypotheses would clearly favour the Organisational approach.

2.3.2 The panel level

The collegial effect of gender

Both the Representational and Informational approach conceptualise gender differences in a rational actors perspective. The first assumes that judges deliberately act as representatives of a particular social group, while the second assumes that judges act upon a particular insight leading them to assume that certain outcomes are better than others are, and to act upon this sentiment. If we assume rationality, we cannot exclude that judges behave tactically, and e.g. do not vote for their preferred option unless they believe they might win through.

This stands in opposition to the Different voice approach, which conceptualise gender differences as fundamental differences between men and women. Although the concept of

tokens has been used in parts of the early Different voice literature to explain the lack of significant gender effects[12], logically we should not expect the effect of gender to be reliant on the amount of women in a panel. For the Organisational approach, assuming no gender effects in the first place, the concept of a critical mass is not very meaningful.

On behalf of the Representational and the Informational approaches, the following hypothesis can therefore be inferred.

H7: "Any individual gender effect will be strengthened in panels with a critical mass – a majority – of women."

The main concept in the Informational approach is the ability judges have to affect each other's behaviour through deliberative forums, which the judicial panel is assumed to be. As such, we should also expect that introducing female members to the court would affect the overall voting pattern of judges. Not only should the general voting pattern be affected by the number of women in a panel, but we should also see an effect of having just one woman (or more) present.

H8: "Controlling for any individual effect of gender, the voting pattern of judges is affected by the gender balance of the panel in which they vote."

H9: "Having at least one woman present will affect the voting behaviour of all judges in the panel."

Because of the strong focus on informational exchange in the Organisational approach, particularities of the deliberative forums in which the exchange takes place is also highly relevant. The presiding judge prepares and leads the deliberation between the judges in the panel. We can therefore imagine that the presiding judge has the possibility to put constraints on – or in other ways affect – the deliberation within the panel. Therefore, a female presiding judge should be assumed to have a better possibility to affect her peers with her own informational background than a female (non-presiding) judge.

[12] See section 2.1.1.

H10: "Controlling for any individual effect of gender, the gender of the presiding judge will affect the voting pattern of other judges in the panel."

As has been touched upon with regard to hypothesis 7, the Different Voice approach does not assume any collegial effects of gender because men and women ration fundamentally different. Similarly, the Organisational approach expects no more gender effects on the panel level than on the individual level. For the Representational approach, the assumptions are fairly similar, although this approach does not exclude that some men can act on behalf of the "female class". This last idea is covered by the concept of critical actors, however. On behalf of these three approaches we infer the following hypothesis, which in effect is the null-hypothesis of the previous hypotheses.

H11: "The voting pattern of judges is not affected by the gender balance in the panel"

The collegial effect of critical actors

Whereas we can assume the existence of critical actors within both the Representational and Informational approach, with regard to collegial effects, the two approaches differ. The Informational approach is the only to assume general collegial effects between gender groups. A logical consequence of this is that, within the Representational approach, we should not expect critical actors to affect their peers in panel, while in the Informational approach, this is a natural consequence of deliberation. On behalf of the Informational approach we can therefore infer the two following hypotheses.

H12: "Parts of any individual gendered voting pattern found in the analysis can be explained by the presence of a critical actor in the panel in which they cast their vote."

H13: "The presence of a critical actor in a panel significantly raises the likelihood of a judge voting in favour of the critical actor's preferred party."

Case particularities

The final group of hypotheses is related to particularities of the case at hand, also measured by variables on the panel level.

Insofar as the Different Voice approach assumes that male and female reasoning differs in a fundamental manner, we can imagine that women are better equipped to convince other

women than other men. Therefore, we can assume that parties represented by female lawyers are more likely to be favoured by the female judges in the panel. Because female lawyers before *Høgsterett* are still the anomaly, we can assume that any differences between judges are strengthened when a female lawyer is present.

H14: "Any individual gender effect is strengthened if the party which female judges are more likely to vote in favour of is represented by a female lawyer"

Because the Representational approach is oriented towards a class conflict between the genders, we can assume that any gender effect found would be strengthened in cases where the female party is pitted against a male party, rather than a representative of the public interest.

H15: "Any individual gender effect will be strengthened when controlling for whether the female party's opposing party is a male individual or a representative of the public."

Finally, inferring from the Informational approach, we should expect the time dimension to affect judicial behaviour. The longer time men have served by women in court, their "informational background" should even out through repeated deliberation. The first year of this selection is set to the year when the first permanent female *Høgsterett* judge was appointed, and for every year after that, we can assume the deliberative effect to even out the differences between male and female judges.

H16: "Any individual gender effect will be alleviated by the number of years women have served alongside men in the court"

3 Methods: A quantitative approach to gendered behaviour

During the last decade, and particularly since 2010, research on judicial behaviour in Norway has gradually been expanded to include advanced statistical analyses of justices' voting pattern in *Høgsterett*. The introduction of more advanced statistical analyses has evolved the field from being limited either to the analysis of judicial argumentation or basic counts of the outcomes of selected cases. Now, more quantitative and variable-oriented analyses are conducted, a development Supreme Court Justice Skoghøy (2011:720) finds fruitful[13]. What is more, with increased sample-sizes and the introduction more advanced methodology, such as multi-level analysis, the study of judicial votes need no longer be limited to non-unanimous panels.

In the social sciences, the general methodological approaches to achieve causal inferences have traditionally been understood in a dichotomous terminology. On one side are the *quantitative* and variable-oriented analyses, normally of a large number of units (so-called *Large-N* studies). On the other side we find the *qualitative* analyses, focusing on either a few or just one unit alone (*Small-N*) (Grønmo, 2004). The debate over which general research design is the better to achieve inference is too extensive for the purpose of this thesis[14]. Rather than attempt an argument for the superiority of a certain general approach, this chapter is devoted to the argument that a quantitative approach, or specifically, a *logistic multi-level analysis* is in itself suitable to answer the research question at hand[15]

The first part of this chapter provides a short overview of the methodological development of the study of judicial behaviour in Norway. The second part of the chapter provides a more thorough description and analysis of the method of choice in in this thesis, *logistic multilevel regression analysis*. Third and finally, an analytical strategy is presented.

[13] Notably, Skoghøy is specifically arguing for the study of non-unanimous panels.
[14] See e.g. Przeworski and Teune (1970), King et al. (1994), Ragin (2004), George and Bennet (2005) or Gerring (2007).
[15] Duly noted, qualitative research designs can be very fruitful in similar analyses. One example is lawyer Ketil Lund (1987), who made use of legal methods to illuminate *Høgsterett's* propensity to favour the public (vs. private) interest in a qualitative study of five controversial cases. MaritSkivenes (2010) conducted a discourse analysis of three cases before *Høgsterett* based on Habermasian deliberation theory. Rather than focusing on how the legal method is implemented, as Lund did, her focus is on the logical consistency of the judges' argumentation. None of these specific methods are, however, suitable to draw causal inference on how male and female judges vote in panels.

3.1 The development of quantitative studies of judicial behaviour in Norway

Norwegian Supreme Court justice Jens Edvin A. Skoghøy (2011) suggests three methodological approaches to the study of judicial behaviour. *The first approach* he suggests is more or less limited to scholars of law and fits into what would be called a qualitative approach, "to analyse the written argumentation for the verdicts" (Skoghøy, 2011:714, my translation).

The second suggested approach is a very basic form of quantitative analysis, the counting of "how large a part of the disputes between citizens and the public are won by the public" (Skoghøy, 2011:718, my translation). The first example Skoghøy (2011) provides of a *quantitative* approach to the study of judicial behaviour is an analysis conducted by then chief justice[16] Terje Wold (1964). Wold literally counts the number of cases won and lost by the state vs. the private party where this dichotomy was relevant. Wold concluded that the variance in outcome was small enough to dismiss any accusations of favouring public parties over private. Studies with fairly similar methodology have been presented by then president of the Norwegian Bar Association, Anders Ryssdal (2006). The latter, however, included all public parties, i.e. also municipalities and counties[17], not just representatives of the state. The latter also found that public parties won more often than private parties did.

Skoghøy lends his preference to the *third proposed approach*, to investigate the votes given by judges in the relevant cases and try to find out whether these votes are connected to the judges' background – i.e. a variable-oriented, large-scale analysis. The prime example of this approach, according Skoghøy, is the study by Tellesbø (2006), building on data from similar analyses by then Supreme Court justice Jan Skåre and Siri Berg-Hansen (1999) and professor of law, Asbjørn Kjønstad (1999). Tellesbø's analysis is a basic count of judicial votes set in connection with factors related to the judges' background. Tellesbø's analysis indicated that the degree to which judges voted either in favour or disfavour of the public party seemed to vary with the judges' social identification.

Arguably, however, the most thorough theoretical and empirical analysis of judicial behaviour conducted by legal scholars is presented by Henry Østlid (1988) study of dissent in his book

[16] «*Høgsterettsjustitiarius*», president of the court.
[17] *Kommune* and *fylke*.

Dommeratferd i dissenssaker. Although the analyses presented in this book are also guided by the same methodological principles as the formerly mentioned analyses, the selection of cases and case fields is much larger than in the former analyses. Among the main implications, Østlid finds that female judges seem to be more lenient in criminal cases[18], and that age, professional background and political attitude seem to affect the dissenting behaviour of judges.

Notably, all the aforementioned legal analyses of judicial behaviour focus on cases of dissent, in which it is reasonable to assume that potential attitudinal differences would be strong. This was also the case in the first analysis of judicial behaviour in *Høgsterett* conducted by scholars of political science. Professors Gunnar Grendstad, William Shaffer and Eric Waltenburg (2011b) improved the methodological state of the field when they made use of logistic regression analysis to study the behaviour of Norwegian Supreme Court judges. With judges' individual vote (in favour or disfavour of the public party) as dependent variable, Grendstad et al. test the predictive power of individual judge's income, gender, former work experience, appointing government and geographical background. The analysis indicates that the latter two variables have a significant effect on judges' voting pattern. The effect of appointing government was confirmed by Grendstad et al. (2012).

Finally, Bentsen (2012), Jacobsen (2012) and Skiple (2012) introduced a further methodological advancement in analysing the behaviour of judges in multilevel regression analyses with the second level of analysis identical with the judicial panels. The main reason behind the fast methodological improvements in the latter years can be found in the development of the "Doranoh" database, initiated and run by Grendstad et al. (2013a). This will be further elaborated upon in the chapter 4.

3.2 The method of choice – logistic multi-level regression analysis

As stated, the chosen research design for this analysis is quantitative. More specifically, the main method of exploring the research question is regression analysis, assuming a two-level hierarchical structure of the data and logistically recoding the dependent variable. This branch of quantitative methods is described in detail e.g. by Rabe-Hesketh and Skrondal (2008), Hox (2010) and Centre for Multilevel Modelling (2013). The present chapter provides discussions

[18] Notably, only three female judges are part of Østlid's analysis.

of methodological problems arising when studying and modelling judicial behaviour and describes how these problems can be accounted for.

The first part draws up a basic outline of how the data should be organised, including a discussion of what exactly is being studied (the dependent variable) and how the explanatory factors ought to be classified. The second part provides a theoretical argument for assuming a multi-level structure of the data when analysing it. The third part, introducing the concept of cross-classification, is a problematisation of the assumed multi-level structure of the data. The fourth part similarly problematizes the data structure, introducing the question of how to account for the temporal dimension. In the fifth part, the procedure of logistic recoding of the dependent variable is briefly introduced, and finally, in the sixth part the analytical strategy is presented in a summarized form.

3.2.1 Modelling judicial behaviour and accounting for dependencies

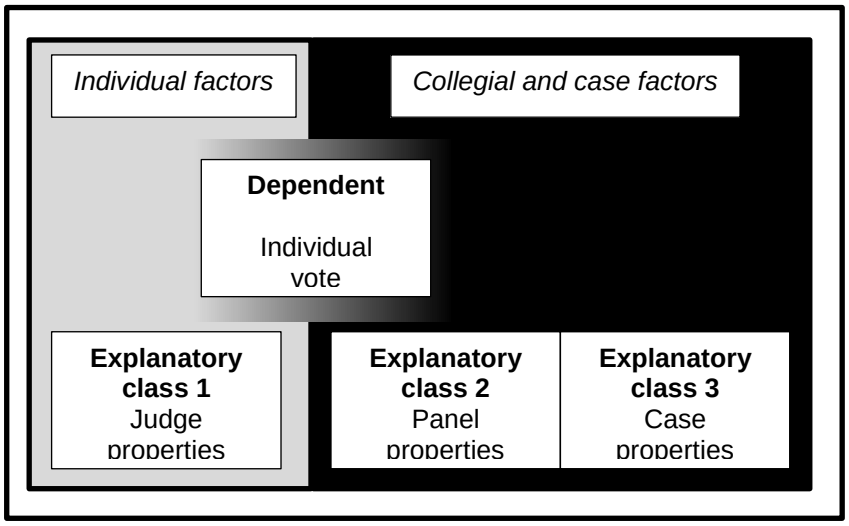
Both ordinary least squares (OLS) and logistic regression analysis assumes absence of *autocorrelation* or *serial correlation*[19], meaning that all observations are independent from each other (Skog, 2009:380, Rabe-Hesketh and Skrondal, 2008:323). Because there are strong theoretical reasons to assume that these assumptions are broken in two distinct manners in the present selection, a multi-level structure of the data is assumed. *First*, judges cast their votes in panels normally consisting of five peers (in this selection, *all* votes are cast in five-judge panels). Since dissent is not the norm, any observation on the dependent variable is likely to provide a fairly efficient prediction of four other observations (all observations in a unanimous panel are equal). We should therefore expect *inter-panel dependencies*, a concept that will be elaborated upon in section 3.2.3. *Second*, the votes are cast by the same judges in different panels, and the chance that judges have a *consistent voting pattern* should be controlled for. Auto-correlation is normal when measuring the same units repeatedly, and is likely to lead to underestimated standard errors, heightening the risk for type 1 errors (Skog, 2009:251-2, Hox, 2010:5). If this is the case in the present selection, a judge's vote in one case can predict the same judge's votes in another case. This will be elaborated upon in section 3.2.4 on *inter-individual dependencies*.

[19] *Autocorrelation* is the most common term for this concept (see e.g. Eikemo and Clausen (2007), Skog (2009) and Bull (2001)), however, in his introduction to econometrics, Baum (2006) uses the term *serial correlation*.

3.2.2 Three classes of independent variables

Since the purpose of this thesis is to analyse to which degree gender affects the behaviour of individuals in the court of last resort in Norway, the dependent variable of this analysis, like in the analyses by Grendstad et al. (2011b), Skiple (2012), Jacobsen (2012) and Bentsen (2012) is the vote of individual judges. As such, when modelling judicial behaviour, we should in principle assume that it is the *votes* themselves and not the *judges* per se who are being studied and analysed. Furthermore, we are trying to predict or at least explain any variance in the votes by different factors or *explanatory variables*. If we accept the premise that it is the votes being analysed and not the judges, there are *only three possible types or classes of explanatory factors* available for the analysis. These are illustrated in Figure 3.1 below.

Figure 3.1: The vote is a factor of both individuals and collegium/case

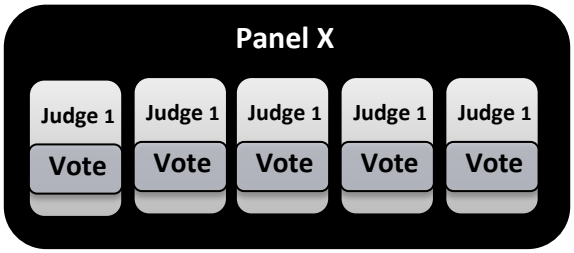


The first class of explanatory factors are *properties of the voting judge*, such as gender, former work experience or geographical background. The second class is *properties of the panel* in which the vote is cast, such as whether there is a majority of women in the panel or whether the chief justice is present. The third and final class is *properties of the case* in question, such as whether the case is a matter of civil or criminal law or whether it is a so-called gender-related issue. This last class would also contain legal factors that are difficult to measure in Large-N analyses, such as how law is applied by the judges. For the purpose of practical statistical modelling, the last two classes are simply assumed to be the same *class* because they are intertwined: When identifying each panel with a unique number, this number equally serves as identification for the case; they perfectly overlap. These two classes of factors will

be referred to combined as *panel level* - or *collegial factors/variables*, while the first class (properties of the judge) will be referred to as *individual factors/variables* (See Figure 3.1).

3.2.3 Inter-panel dependencies – votes are cast within panels

Figure 3.2: Votes are cast within panels



An important property of votes cast in *Høgsterett* is that they are cast within panels of judges[20]. A property of these panels is that they are deliberative, which is the feature Skivenes (2010) made proper use of in her qualitative study of the deliberative quality of three child custody cases. Another important property is that the deliberations of a panel of judges always have to end up in a majority-supported conclusion[21]. This feature is particular to the justice system in comparison with many other deliberative forums of decision-making, e.g. parliament, where conclusions can be limitlessly postponed if the cases at hand are too hard to handle. In the words of the judges themselves, invariably more than 50% of the judges in absolutely all panels agree with each other "*in the essentials and the result*"[22]. We should therefore assume that the mean level of agreement within the panels of judges is very high, and when modelling the variance in votes, we need somehow to account for this structure. Thus, we assume that votes are *clustered* within panels.

[20] Most verdicts cast in *Høgsterett* – and all verdicts in the present selection – are cast in five-judge panels. The two alternative panel sizes are the *grand chamber* ("Storkammer") of eleven judges and the *plenary session* ("*Plenum*") of all members of the judicial collegium, giving an indication of the perceived importance of the cases in question. In addition, there is the three-judge *appeals chamber*, only evaluating whether to allow appeals. The fact that there are only five-judge panels in this particular selection is coincidental.

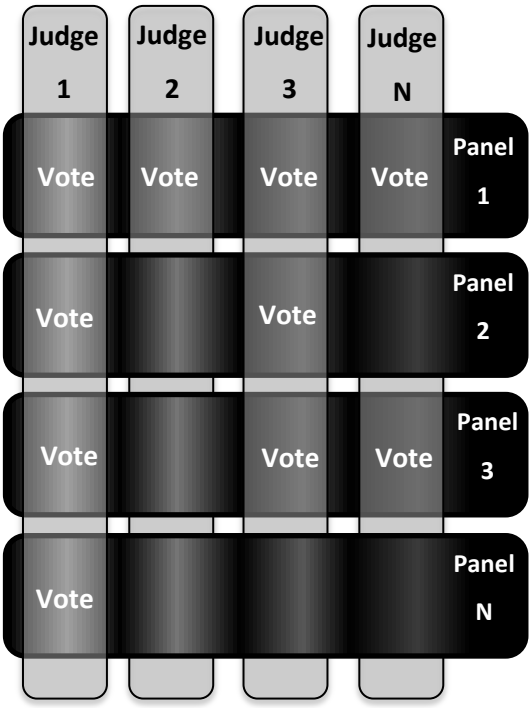
[21] Notably, although the judges are bound to reach a conclusion in the case at hand, the precedent from a judgement can still be vague and unclear, comparable to the outcome of a parliamentary debate in which majority is hard to achieve. The matter at hand here is that *some* conclusion will get majority support, not whether the quality of this conclusion is good or bad.

[22] The reports of the judgements in *Høgsterett* are written by *the first voting judge* (*førstvoterende*). If they do not dissent, the remaining four judges' consent is represented by the formalised phrase "I, in the essentials and the result, agree with the first voter" ("*Jeg er i det vesentlige og i resultatet enig med førstvoterende"*).

3.2.4 Inter-individual dependencies and cross-classification

Accepting the premise that the object of study is the votes themselves and not the judges, in addition to accounting for *inter-panel dependencies*, we also need to account for *inter-individual dependencies*. As such, it would be reasonable to assume that there are three levels to the analysis. We would assume that votes are cast "within" judges, who work within panels. Such a three-level hierarchical model is exemplified in the leftmost classification diagram in Figure 3.4. Votes are cast by – *or clustered within* – judges, who in turn work in – or *are clustered in* – panels.

Figure 3.3: Votes are cast by individuals in different panels with changing constellations

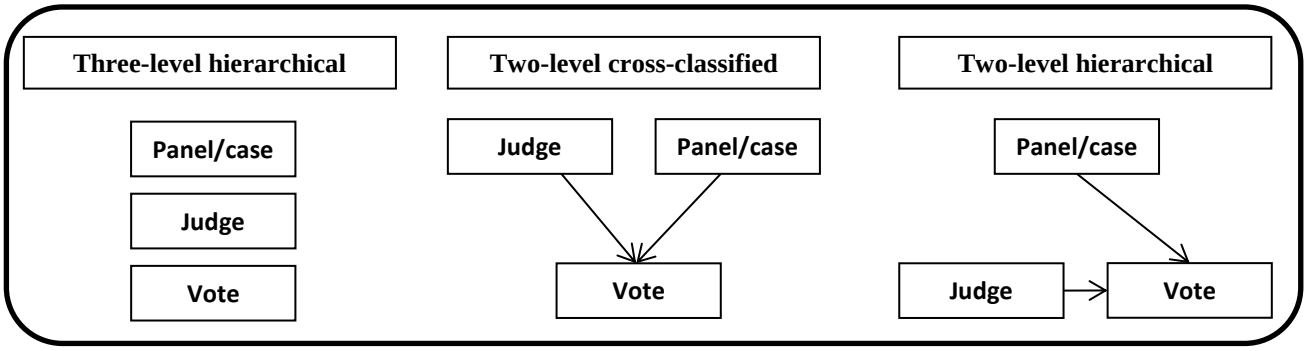


However, this only works until we add more than one panel to the model. The problem arising when studying several panels is illustrated in Figure 3.3. For every panel of judges drawn in *Høgsterett*, the choice of judges to take part in the panel is conducted in a near arbitrary manner[23]. A necessary result of this is that the composition of panels varies from case to case. As such, the properties of judges cannot be assumed to be uniformly and hierarchically positioned *within* panels. Rather, we see in Figure 3.3 that the two classes of explanatory factors form a pattern similar to that of woven fabric. This is a typical pattern of a *cross-classified* data structure, as illustrated in the middle classification diagram in Figure 3.4. The variance in the dependent variable is indeed explained by two different *classes of explanatory factors*, but these classes are not hierarchically related to each other[24]. Hox (2010:172-3) suggests that it makes sense, in an example similar to the present data pattern, to view the data as a cross-classification where factors related to the *measurement occasions* (i.e. panels) *individual subjects* (i.e. judges) are both treated as second level variables without any direct hierarchical relation to each other.

[23] For a detailed review of the process of allocating judges to panels, the so-called *controlled lottery*, see Grendstad et al. (2013b)

[24] See e.g. Leckie (2013:8) or Fielding and Goldstein (2006), see Stata output in the appendix, section 7.1, for reference.

Figure 3.4: Three potential multi-level structures



Assuming a cross-classified, multi-level structure of the data, however, leads to large and complicated models, which in turn strains both hard- and software much more than a simpler hierarchical model (Hox, 2010:185-7). Attempting to run these analysis with the available computer power as proven to be futile.

The common practice in multi-level analyses of judicial behaviour in *Høgsterett* has been to model the dependent variable (the individual vote) as *properties of the judge*[25], and assuming a two-level hierarchical structure of the data. Thus, the cross-classified nature of the data is ignored. This solution has been chosen in the present thesis as well, however doing so still leaves a need to somehow control for potential *intra-individual* dependencies in votes cast by the same judges.

Critical actors as a measure of inter-individual dependencies

As has been touched upon earlier in this chapter, exploring the existence and potential effects of *critical actors* in this thesis has two purposes. The *first purpose* is the obvious one conceptualised in the theory chapter (section 2.2), to identify whether critical actors can be said to exist in the selected data and to analyse whether these actors have any effect on their peers' voting behaviour. The *second purpose*, however, is methodological. The *critical actor* analysis serves to control for *inter-individual dependencies*, as described in section 3.2.1. Step 2 of the analytical strategy is centred on the identification of critical actors in the selection, while in step 5, the general effect of gender is controlled for the effect of any critical actors. See sections 3.3.4 and 3.3.5 under *Analytical strategy* for an elaboration.

[25] See Bentsen (2012), Skiple (2012) and Jacobsen (2012).

3.2.5 The temporal dimension

Thus far, the temporal dimension has been left out of the discussion. Data has been collected from 1968[26] to 2011, which is a considerable span in time. The starting year, in addition to being the year when Lily Bølviken was appointed first female judge on the *Høgsterett* bench, also connotes drastic attitudinal change. This was a year of social tumults, which – in hindsight – seems to have forespoken drastic changes in the common perception of gender, equality, sex, etc. In the following decades, women have gained representation in executive position formerly understood as part solely of the male domain. A somewhat similar change has happened in men's position in social spheres formerly understood to be *female*. In 1968, the idea that fathers would be allocated a particular *father's quota* of paid parental leave would probably seem absurd to most Norwegians, while this – in 2011 – was the reality. This anecdote does not provide any evidence, but serves to illustrate that we ought to assume that the popular opinions on child care and gender roles might have changed radically from the start of this period to its end. Equally, we should assume that judges are equally affected by such general attitudinal changes as anyone else living in Norway. We should thus somehow control for this temporal dimension when analysing gendered judicial behaviour. To do so, two time variables are added as control variables to provide a crude measure of the temporal dimension. One of these is a metric variable where the distance between each value is one year. The other is a dichotomous dummy variable, approximately splitting the selection in two groups over the temporal dimension.

3.2.6 Logistic transformation of the dependent variable

Because the dependent variable proposed for this analysis is dichotomous, the dependent variable is logistically transformed[27] to avoid breaking the assumptions of linearity, homoscedastic errors and normality (Hox, 2010:112). Specifically, this is needed to make estimation of the model possible. In addition, it counters false estimations of the coefficients because of non-linearity (Skog, 2009:238) and false estimations of the standard errors and thus p-values because of heteroscedasticity (Skog, 2009:246-7) and non-normality (Skog, 2009:249).

[26] This is the year Lilly Bølviken was appointed the first female judge on the *Høgsterett* bench.
[27] See Skog (2009:354-8 & 240-6) for a thorough explanation of this process.

3.2.7 Additional assumptions

At every step of the analysis, the data has been tested for *multicollinearity*, which is particularly important with variables measuring similar or the same phenomena. Chen et al (2003:chap2) recommends running collinearity diagnostics using the command *collin* in Stata, presenting the variance inflation factor (VIF) also recommended by Acock (2010:262-3) and Eikemo and Clausen (2007:126-7). In addition, all variables included in the analysis have large enough amounts of observations to avoid indications of *discrimination* problems, as these are described by Eikemo and Clausen (2007:129-30).

3.2.8 In summary: A model of gendered behaviour in the court room

The present analysis assumes a two-level hierarchical structure of the data, with judges and their votes on the first level – referred to as the *individual level*. These observations are clustered in cases or panels on the second level – referred to as the *collegial level*. This chosen structure is equal to the two-level hierarchical model presented in Figure 3.4 on page 29.

This structure, clustering the data only in panels and not in judges, builds on the assumption that votes are properties of the judge alone. In fact, however, it is reasonable to assume that they are a function of factors related to both the judge, and the panel (both related to the composition of the panel and the particularities of the case at hand). As discussed, this is a pragmatic choice to avoid large and complicated models, although it does ignore the cross-classified nature of the data.

To alleviate this problem, the analysis of potential *critical actors* is given a methodological function in addition to being of empirical interest in itself. Adding controls to the model for those judges who stand out as critical actors, also serves as a control for potential *intra-individual dependencies*.

Finally, the temporal dimension is controlled for, by two control variables measuring time.

3.3 Analytical strategy

To recapitulate, the research question for this thesis is "*How and why does gender affect the judges of the Supreme Court of Norway?*" The first two steps of the analysis are directly linked to this question: In the *first step*, an empty, unrestrained model is run to calculate the intra-class correlation. In the *second step,* the general effect of gender is analysed, first on the individual level, then on the collegial level. This part of the analytical process is theory testing and deductive and results in a temporary model of the significant effects on individual judges' voting pattern in cases of child custody.

In the *third step*, however, the analysis crosses over into an explorative and inductive subject. This step is related to the identification of potential critical actors, as argued for in the theory chapter, section 2.2 and the methods chapter, section 3.2.4. This process leads to a list of judges whose voting pattern is significantly consistent throughout the cases in this selection.

In the *fourth step*, the deductive and the inductive processes are combined, as the effect of critical actors is added as control variables to the final model in step two.

The general analytical strategy in all steps of the process, however, follow the recommendations given by Hox (2010:56-59): The analyses are run block-wise, starting with an empty model[28] containing no explanatory variables, then gradually adding variables, first on the individual level, then on the collegial level. Multi-level models are quite complex not only with regard to the number of parameters, but also in terms of interpretation (Hox, 2010:55). This general strategy of adding one and one variable makes it possible to evaluate whether to keep or abandon each variable at every step of the process, thus keeping the model as parsimonious as possible.

3.3.1 Hypothesis testing and model comparison

For hypothesis testing two different statistical tests will be applied[29]. *First*, for testing the null-hypothesis for individual explanatory variables the z-ratio ($\hat{\beta}/SE(\hat{\beta})$) of each variable's

[28] This is also variably referred to as an *unconstrained model* or *intercept-only model*.

[29] Note that a third method of hypothesis testing, the Wald test, could also have been applied to test the null-hypothesis for entire blocks of explanatory variables. This type of significance testing is especially useful where one categorical explanatory variable is represented by a set of dummy variables (Steele, 2009a:30). This is not the case in any part of this analysis. As such, the Wald test would not provide any information significantly different from what we get from the LR test. See also Steele (2009a:31).

estimated coefficient is compared to a standard normal distribution (Steele, 2009a:29-30). Noteworthy, this test statistic (p-value) procedure by this procedure is a function of both the coefficient and its standard error, meaning it can easily be affected by broken assumptions that affect predicted standard errors, such as multicollinearity (see e.g. Skog (2009:286-288) or Field (2009:224)).

Second, a *likelihood ratio (LR) tests* is recommended by Hox (2010) to measure the change in goodness of fit between *nested* blocks in multilevel analyses. The LR test is analogous to the F-test used to compare linear (non-logistic) regression models, and conveniently produces a test statistic (p-value), which can be interpreted as a measure of whether there is a significant increase in explanatory power between the two blocks. See e.g. Hox (2010:47-50) or Steele (2009a:30-31).

In addition, two different *goodness-of-fit* measures will be presented for each block of the analysis. Although not intrinsically interpretable (Field, 2009:781), *information criteria*, such as Akaike's information criterion (AIC) and Schwartz's Bayesian information criterion (BIC) provide useful statistics for comparing different models' goodness of fit. Low values of AIC and BIC[30] indicate good fit in the sense that estimated values are close to observed values (Field, 2009:304). This is a practical tool for comparing *non-nested* models and can be used to compare model fitness between the different empty models. AIC/BIC will also be calculated in the following steps of the process and used as a reference, together with the LR-test (see step 2 of the analytical strategy), to compare model fitness.

3.3.2 Step 1: The empty, unconstrained models

As argued for in sections 3.2.3 and 3.2.4, there are strong theoretical reasons to assume *inter-panel* and *inter-individual* dependencies. These assumptions can be tested by calculating the *intra-class correlation* (ICC) in empty or unconstrained models[31] where the dependent

[30] It is normal to include both information criteria, although they normally provide similar estimates. While BIC is slightly more conservative than AIC, both are corrected for model complexity by adjusting for the number of parameters estimated in the model (Field, 2009:781-2). An additional problem with BIC particular for multi-level models is that it relies on N, but it is unclear which level we should retrieve N from. Here, level 1 N is being used, which is normal. See a discussion of BIC in StataCorp (2011: 157-161).
[31] See footnote 28.

variable is grouped around either *judges* or *panels*. This is the first step of the analysis, found in section 5.2.1.

The intra-class correlation

In two-level regression analyses, we define a second level, or a "class" (e.g. the panels which judges vote within) and a first level (e.g. the judges themselves and their votes, see Figure 3.4). The *intra-class correlation* (ICC)[32] is calculated by dividing the level-2 (residual) variance in the dependent variable by the total variance in the dependent variable. This provides a standardised coefficient, which can be interpreted as a percentage measure of how much of the total variance in the dependent variable happens between, rather than within, any group or class that we have defined. (Leckie, 2010a:7).

If there is as good as no *within-group variance*, e.g. all judges always vote the same as their peers in panel, the ICC would be very close to 1. This indicates that a large part of the explanation of how the outcome in the dependent variable varies is to be found in differences between the level-2 units (e.g. panels), rather than between the level-1 units (e.g. judges). If, on the other hand, the *within-group variance* were high, the ICC would be closer to zero. This would indicate that a large part of the difference is to be found in differences between the level-1 units of the analysis, rather than the level-2 units.

As such, it can be said to measure the level of autocorrelation in the dependent variable *within* certain "classes" or groups of observations. When modelling a dichotomous outcome, such as whether judges vote for or against a feminine party grouped or nested within panels, we can therefore interpret the ICC as a measure of the level of agreement within these panels. The basic interpretation is that if the ICC is high, there is a high level of agreement, thus little dissent in the panels. If there is a high level of agreement, a large part of the explanatory factors ought to be sought in differences between the panels (related to panel composition or case particularities, see section 3.2.2). If we model the votes grouped or nested within judges, we can interpret the ICC as a measure of the level of *judge self-agreement*, i.e. whether the

[32] Steele (2009b:16-17) more precisely refers to this as the *Variance Partition Coefficient* (VPC). The VPC is equal to the ICC in simple multi-level models (Steele, 2008:8) and is commonly referred to as the ICC. Both VPC and ICC will therefore be referred to as ICC in this thesis. The formula is $\frac{\widehat{\sigma_u^2}}{\widehat{\sigma_u^2}+\widehat{\sigma_e^2}}$, where $\widehat{\sigma_u^2}$ is the level-2 residual variance and $\widehat{\sigma_e^2}$ is the level-1 residual variance (Steele, 2008:8).

overall judge votes similarly over different cases. Which type of model best serves this analysis is determined in the first step of the analysis, in section 5.2.1.

An important property of the ICC in logistic analyses is that the level 1 residual variance is fixed at 3.29 as a result of the logistic recoding of the outcome (see Steele (2009b:17)). As such, the intra-class correlation in logistic analyses is mainly affected by changes in the second level variance term, whereas it "normally" would rely on variance terms from both levels (Hox, 2010:59). In the present analysis, we therefore expect the ICC to remain approximately unchanged when we add a variable that explains level-1 variance. If we add a variable that explains (and thus controls for) level-2 variance, we expect it to decrease.

Panels

According to (Acock, 2010:237-9), the ICC can be interpreted as a percentage measure of agreement within groups. Therefore, if the ICC is high in an empty model where the votes are grouped by the panels in which they are cast, this is interpreted as a high mean level of agreement within the panels of judges, indicating that we need to look to the collegial or case level to explain the main variance in the dependent variable.

Individuals

Similarly, a high ICC in an empty model where the votes are grouped according to the voting judges is interpreted as a high level of mean "agreement" between votes given by specific judges. Simplified, this shows that the vote given by an individual judge in one case is similar to votes given by the same judge in other cases. The practical interpretation of this is the degree to which the mean judge gives the same vote (in either favour or disfavour of the feminine party) in subsequent cases.

Cases of dissent

For the reference, it is also interesting to calculate the ICC including *only cases of dissent*. Given that the ICC is a measure of agreement within panels, we should assume a much lower ICC in this part of the selection. In addition, to have a measure of the general degree of *intra-individual dependency*, we calculate the ICC of an empty model in which the judges are assumed to be the second level of analysis

3.3.3 Step 2: The general effect of gender

Gender on the individual level

In step 2, level 1 variables are added block-wise to the model in accordance with the general strategy recommended by Hox (2010) and presented in the first paragraph of the analytical strategy.

Gender on the collegial level

When adding variables on the second level of analysis, in addition to the *LR-test* and the *AIC*, we can expect the ICC to provide interesting information on the comparison of blocks. The ICC is a function of the level 2 residual variance divided by the total residual variance (both levels). Since the dependent variable has been recoded with a fixed logistic normal distribution, the level 1 residual variance is also fixed[33] and contributes little information. As such, we should expect little or no change when adding variables on the first level of analysis. When controlling for a significant variable on the second level, however, we should expect the residual variance on this level to go down and consequently also see a reduced ICC, leaving a larger share to be explained by the first level variables.

Those variables found to have a significant effect on the dependent variable are then included in a new model which is compared through an *LR-test* to the end model of the previous step. This procedure is repeated for all following steps.

3.3.4 Step 3: Identifying critical actors

The process of identifying critical actors is elaborate and starts with the identification of critical individual judges through dummy variable coding. *First*, every judge appearing at least 10 times in the selected panels is identified by a dummy variable. For example, any observation where chief justice Tore Schei has given a vote, the dummy variable "*Schei*" is coded 1. All other observations on this variable are coded 0. *Second*, we wish to measure any potential effect of the critical actors on their peers. To do this, a dummy variable is coded 1 where the judge is present in the panel and coded 0 in those panels where the judge did not partake. Referring the former example, in *all five observations* in any panel in which chief justice Schei took part, the dummy variable "*Schei in panel*" is coded 1. All other

[33] The ICC is calculated by $\frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2}$, and the level 1 residual variance (σ_e^2) is fixed at 3.29 ($\pi^2/3$). For further explanation, refer to Steele (2009b:7.2) or Rabe-Hesketh and Skrondal (2008:256-7).

observations on this variable are coded 0. In the terminology of Hox (2010:2-4), the first kind of variable described here is a *global variable*, referring only to its own level of analysis, while the second kind is a *structural variable*, a variable constructed by aggregating information from a lower to a higher level of analysis.

These dummy variables' potential of predicting the outcome of the dependent variable is then tested first in bivariate models and then trivariate models[34]. *The bivariate models* predict the outcome only based on the individual-identifying variable. If this variable is found to significantly[35] predict the outcome in the dependent variable, this is interpreted as a *consistent voting pattern*, i.e. that the judge has been voting fairly consistently either for or against the feminine party (depending on the direction indicated by the coefficient). This is the criteria set for identifying a judge as a potential *critical actor*.

The next step, to analyse whether the critical actors affect their peers is conducted in *the trivariate models*. These are similar to the bivariate models, but also include the aggregate variable measuring the judge's presence in a panel. Notably, the individual-identifying variable is always included to avoid measuring a judge's effect "on herself". Technically, once again referring the example of chief justice Schei, by including the individual-identifying variable "*Schei*" while testing the aggregate variable "*Schei in the panel*", we keep the Schei-variable constant while measuring the effect of Schei on his peers. If this panel variable is found to predict the outcome significantly, this is interpreted as showing that the judges who have been serving in the same panel as the judge in question have been voting consistently in the direction indicated by the coefficient. If the direction is similar to that of the individual-identifying variable, this is interpreted as an indication that the judge in question has affected his peers' vote.

This process is too elaborate to be included in the analysis chapter, where only a table including the effect of those judges identified as potential critical actors are included. However, a full set of analyses is included in the appendix.

[34] These are, of course, also two-level hierarchical models, assuming properties of the judges to belong to level 1 of the analysis and properties of the panel/case to belong to level 2.
[35] On a significance level of five per cent.

3.3.5 Step 4: Adding the effect of critical actors on the gender model

In step 4, the general effect of gender is controlled for the effect of those judges identified as critical actors in step 2, thus intertwining the two main parts of the analysis.

When comparing the goodness of fit of non-nested models, Hox (2010:50-1) recommends a comparison of AIC[36] for the two models. Hox suggests that, in comparison between two models we should choose the most parsimonious. The AIC tends to be reduced when the deviance goes down, which indicates a better fit, thus the model with the lowest AIC is preferred.

3.3.6 Step 5: Varying slopes

At all former steps of the process, we have allowed only the intercept to vary between panels. Following Hox's the suggested analytical strategy, the next natural step in the process is to measure whether any of the explanatory variables' slopes vary significantly between the panels (Hox, 2010:58). If the effect of gender differs from panel to panel, this could be identified at this stage of the process. Notably, since only the average slope of the individual-level variables omitted in step 3 have been tested, these can be analysed again at this stage.

The danger of over-parameterisation, leading to various estimation problems is imminent at this stage. Thus, explanatory variables should be tested for varying slopes one by one. Those variables found to have significantly varying slopes are then included in a new model which is compared to the final model of the previous step through an LR-test

3.3.7 Step 6: Cross-level interactions

If significantly varying slopes are identified in the previous step, we should try to identify why the effect varies from panel to panel. Referring the example from the previous step, if the effect of gender is found to differ between panels, this is where we can identify which properties of the panel affects the slope variation, thus giving us a measure of the effect of gender within different ideal-typical panels. There are no specific theoretical reasons to assume cross-level interactions in this analysis.

[36] *Akaike's Information Criterion*. This is preferred over Schwartz's Bayesian IC because of the ambiguity related to calculation of the latter which is a function of (among others) N. In e.g. a two-level model, however, it is unclear whether N refers to the first or second level. In practice, however the two perform equally well, but with a slight advantage for BIC (Hox, 2010:51).

4 Data

The data collection for this thesis has been a very demanding job. First, there was a need to operationalise the concept *gendered issue* in the Norwegian context, where many of the issues known from e.g. the US and Canadian courts are less frequently heard in court. In the following, a lot of work was laid down in finding the relevant information through the legal database *Lovdata*, and developing search criteria that would find all relevant cases within the operational definitions of the case categories. Finally, the data had to be manually copied into the Doranoh database structure[37].

This chapter presents the data collection process and the data that has been analysed. First, the main data sources are introduced. Second, the collection process is elaborated upon from contextualisation of *gendered issues* to the collection and quantification of the data. Third, potential sources of bias in the data are discussed. Finally, the variables in the analysis are presented and given operational definitions.

4.1 Data sources

There are two main sources to the data that forms the basis for this thesis. One source is the Norwegian legal database *Lovdata*[38], in which all judicial decisions and votes in the Norwegian justice system is digitalised and made publically available. The second source is the *Doranoh*[39] database (Grendstad et al., 2013a). Some background information has also been gathered on individual judges where this has been lacking in *Doranoh*. The main source of this has been a range of Norwegian newspapers, through the digital newspaper archive service *Retriever*.

Doranoh has provided two kinds of data otherwise unavailable. The first is voting information in cases that, for different reasons, have been left out of *Lovdata*, generally referred to as *star cases*[40]. The second is information about the individual justices of *Høgsterett*, making possible the analysis of how properties of individual judges affect their voting pattern.

[37] Already working as a research assistant for the Doranoh project was very helpful in this respect.
[38] For detailed information, see the English Lovdata web page: http://lovdata.no/info/lawdata.html
[39] The data in the *Doranoh* database ("*DOmmerAtferd i NOrges Høyesterett"*) has been collected through the research project *Judicial Behaviour in the Supreme Court of Norway*, led by professors Gunnar Grendstad, William R. Shaffer and Eric N. Waltenburg. The undersigned to this thesis has taken part in extensive data collection for this project as a research assistant from 2011 to 2012 – in addition to working on this thesis.
[40] The name refers to the (star shaped) asterisk added to the end of these cases' reference number in *Lovdata*.

In addition to providing data, the flexible design of *Doranoh* has provided the infrastructure necessary for advanced quantitative research on judicial behaviour in *Høgsterett*. As such, all data collected from the qualitative *Lovdata* database has been quantified through the *Doranoh* infrastructure, and is stored in this database both for future research and for the purpose of data reliability. Without this infrastructure, the analyses in this thesis would have been very hard to conduct.

Quite a lot of data had to be collected for the analyses in this thesis. Because *Doranoh* is a relatively new database, some categories of information are somewhat underdeveloped. Particular for these analyses are the lack of data on consensus panels and the lack of data on the parties meeting before the court. For the cases included in this thesis, both of these lacunae have been filled manually by registering qualitative data from the *Lovdata* database to the *Doranoh* infrastructure.

Fundamental to adapting the theoretical framework presented in the theory chapter to a Norwegian-specific context has been the selection of an area of law to focus the analysis on. Gendered case categories like e.g. *abortion* or *discrimination in the work place* are less dominant in the Norwegian court system than in e.g. the US American courts. The process of developing an operational case category for a relevant gendered issue has also taken a lot of resources.

The first purpose of this data section is to give an account of the data collection process by elaborating on the issues mentioned above. The second purpose is to present the operationalization of the concepts, case categories and variables forming the basis of analysis.

4.2 Data collection

4.2.1 Phase one: Contextualising the concept of *gender-related issues*

Although gendered judicial behaviour has never been the main focus of any analysis of the Norwegian Supreme Court, a few articles have brought footnoted hints and the like, as to where gender effects potentially could be found (see e.g. Drammens tidende (2009) or Skivenes (2010)). This was combined with a run-through of the "main register" of *Norsk retstidende* before and after 1996, to get an impression of which categories the legal community operates with when defining an area of law and how they operationalize these categories. From this process emanated a list of six categories of cases which might both be

assumed to be *gender-related* and which could allow for statistical analysis with testable hypotheses:

1. Cases related to the child care service[41]
2. Criminal trials related to sex[42] motivated crimes
3. Criminal trials with young defendants
4. Cases related to confirmation of fatherhood
5. Cases related to the freedom of speech
6. Cases related to pornography

4.2.2 Phase two: Exploring the possibilities for analysis

These six case category suggestions formed the basis for an iterative and time-consuming process of searching the *Lovdata* database for any relevant cases. Similar to the first phase, this part of the data collection was explorative, and the criterions for which cases to include and exclude were set and revised several times *during* and as a part of the search. The purpose of this was three-fold:

a) to collect exhaustive lists of all cases in the selected time period that fit within the selected six case categories;
b) to be familiarised with the selected areas of law, and;
c) to evaluate the extent to which these categories held enough cases to allow for a solid multi-level regression analysis.

As presented earlier in this chapter, the purpose of *Lovdata* is to be a judicial archive for practising lawyers, and not at a source for rigorous analysis. Making a selection of cases through *Lovdata*, while avoiding any sort of incidental selection bias is therefore a challenge. The selection was therefore conducted in three steps. *First*, complex search criterions were constructed on the basis of relevant keywords for each area of law. These were then used for automatic searches through the case abstracts[43] in *Lovdata*. *Second*, the case abstracts in *Lovdata* have been scanned for relevant references to the law[44]. *Third* and finally, every case

[41] *Barnevernet*
[42] Sex as in sexual relations, not gender.
[43] The abstract in *Lovdata* are manually written when the cases are added to *Lovdata* from *Norsk retstidende*.
[44] When documents are added to *Lovdata*, legal references are automatically extracted from the transcribed legal argumentation of the judges and added to the case abstract.

in the automatic selection have been manually controlled and read through to exclude all cases that prove to be irrelevant.

Three of the case categories were abandoned early on in the process due to difficulties of data collection or small number of cases considered by *Høgsterett* in the selected period. Specifically, while developing efficient search criteria to find all criminal cases involving *young defendants* proved difficult. the number of cases involving the *freedom of speech* is low[45] and the number of *pornography* cases even lower in the period. The three remaining categories are deemed fit for quantitative analysis:

1. Cases of sexually motivated crime
2. Cases related to confirmation of fatherhood
3. Cases of disputed child custody[46]

Because of the relative difference in nature between these three case fields, coding a common dependent variable to include all in the same analysis is difficult. As such, to avoid having to run three completely separate analyses, the case selection was narrowed down to only one category: *cases of disputed child custody*.

Although labour-intensive, the double search for both references to manually written keywords and to law references and subsequent manual proofreading should guarantee that the case selection includes an exhaustive list of cases within the chosen category. Thus, the case selection in this thesis does not discriminate between cases perceived to be more or less interesting in a legal perspective. In addition, the selection makes out *the entire universe* of child custody cases between 1987 and 2011.

4.2.3 Phase three: Collecting and quantifying the case information

Having a list of cases is not enough to run an analysis. The third phase of the process is collection and quantification of the qualitative data from *Lovdata* into the *Doranoh* database.

[45] This is despite a the presence of a few high profile cases, such as the *White Electoral Alliance* case (Hvit valgallianse), *Rt-1997-1821*

[46] Note that this case category is an expansion of the first category in the first selection, which just included cases related to the childcare services. By also including disputes between parents, this selection now includes all cases of dispute over whom the care of a child should befall.

While voting data was added where this was missing[47], all the data on the parties and their lawyers have been collected for this thesis specifically. Some background information has also been gathered on individual judges where this has been lacking in *Doranoh*.

The next step of the process is the operationalisation of variables and concepts, which will be elaborated upon in section 4.4, after a presentation of potential sources of bias related to the case selection.

4.3 Potential sources of bias

4.3.1 Potential bias in case abstracts

All searches in Lovdata have been limited to the case *abstracts*. These abstracts are manually written, not by the judges but partly by the *Norsk retstidende* editorial staff and partly by the *Lovdata* staff, and thus also of varying quality. The thorough data extraction process is meant to ensure an exhaustive extraction of cases within the chosen case category, but can of course not guarantee against lacking information in case abstracts. This potential bias is attempted alleviated by an operationalisation of the case field which relies on both the automatically collected law references[48] and the manually written abstracts.

4.3.2 Consensus and dissent

Much of the focus of scholars of judicial behaviour has been put on cases of dissent. Some scholars argue that the primary objects to study judicial behaviour in is cases of dissent or non-unanimous cases. The fact that the justices are in dissent is assumed to indicate that the formalised legal sources do not provide clear cut answers, thereby making the justices' individual preferences more prominent in explaining the cast votes (see e.g. Grendstad et al. (2011a:20) or Østlid (1988)). Focusing merely on dissent cases was also argued to be a way of *objectivising* analyses by Norwegian Supreme Court justice Jan Skåre (1997). In addition, when conducting large-N analyses, this is a way of saving resources related to data collection, since unanimous cases make out the absolute majority of cases before *Høgsterett*. Thus, because little data on cases of consensus (non-unanimous cases) in *Høgsterett* has been available through Doranoh[49], dissent cases alone have been the focus of research. Focusing merely on a small part of the available data can be problematic, however.

[47] Particularly for unanimous cases.
[48] *Lovdata* automatically collects law references from the published legal arguments.
[49] The database was expanded in 2011, see Bergset (2011).

When studying gender effects in judicial panels, Peresie (2005:1764-5) points out that non-unanimous cases by definition involve disagreement. As such, any theoretical approach to gendered behaviour which assume gender differences, must also assume that justices are also more likely to disagree over gender lines in non-unanimous cases.

Say, for example, that no effects are found when modelling the judicial vote on an aggregate of the gender variable (a measure of the gender balance in the panels). This reason for this might be that when judicial panels are divided, male and female judges tend to end up on opposite sides, as is the case both in Skivenes' (2010) small-N selection and in Kjønstad's (1999:101-2) argument for potential gendered behaviour. Even though this was the case, we still had no reason to assume that female and male judges were not perfectly able to compromise in unanimous cases, where perhaps the number of female judges in panel might affect the voting pattern of male judges (or the other way around). Seemingly, if gender differences prove to be strong in the following analysis, including only cases of dissent, could lead to a faulty conclusion that gender only can explain variation *within* the panels, and not *between* the panels. This is similar to what King et al. (1994:129-137) refers to as *selection on the dependent variable*.

As such, excluding cases of consensus means excluding a large selection of cases where an aggregate panel effect of gender might explain more of the gendered voting behaviour of judges. This is particularly a concern for the *Informational* approach to gendered behaviour, in which we need to assume that judges affect their peers' attitudes through deliberation in the panels. When studying panel effects, cases of consent are particularly interesting, as it is interesting to observe whether unanimous judicial decisions can be the result of other factors than merely legal factors.

Despite the extra cost of data collection, based on the above methodological considerations both cases of consensus and dissent are included in the present analysis.

4.3.3 Limited to a single case field

Another potential bias, the narrow selection of cases in the present thesis, is similar to what King et al. (1994:137-8) refers to as *selection on an explanatory variable*. The case field is chosen *because* it is a *gendered issue* and *because* there are theoretical reasons to assume that

gender effects can be observed in these kinds of cases. As such, the analysis is searching for gendered differences only where differences are likely to be found. However, since this selection procedure does not predetermine the outcome of the analysis, this does not cause any inference problems, according to King et al.

In addition, Peresie (2005:1765) argues that several analyses of gendered judicial behaviour are flawed because they include a broad range of case types, rather than focusing on particular case fields. It would be unreasonable to assume that legal factors, such as the sources of law, did not make out a large part of the explanatory factors for judicial behaviour in *Høgsterett* (see e.g. Epstein and Knight (2013), Sunde (2012b) or Bentsen (2012)). As such, it is reasonable to assume that potential individual effects, i.e. gendered effects, are reliant upon particularities of the cases being studied, such as the case field. By including too broad a range of cases and not controlling for the impact of case particularities, gendered effects restricted to specific case fields (such as cases of child custody) are minimised.

That said, within the theoretical framework of this thesis, the *Different voice* approach assumes gendered effects to be found within all types of cases, while the other two models restrict these effects to *gender-related issues*. An analysis including – and controlling for – both *gendered* and *non-gendered* issues would therefore strengthen the validity of the results. The choice to, despite this, narrow the selection down to a single case field, is a matter of pragmatism due to the heavy workload associated with the collection of data.

4.3.4 Potential bias: Lack of important control variables

Peresie (2005:1765) argues that the analyses of Farhang and Wawro (2004) and Massie et al. (2002) are empirically flawed in that they have failed to control for enough individual characteristics of the justices. This results in potential over-estimation of gender effects, or all-out spurious effects (Grønmo, 2004:295, 366-8, Skog, 2009:37-45). She recommends including control variables representing judges' past careers, age, and federal appellate experience in addition to ideology and race. In the Norwegian context, race would more or less be a constant. The remaining four factors are included in the analysis, however. This can never serve as a guarantee against spurious effects, however.

4.4 Operational definitions of variables

4.4.1 Dependent and explanatory variables

Table 4.1: Operationalisation of the dependent and the explanatory variable(s)

Theoretical concept	Variable	Operationalization
Dependent variable		
Dependent variable	**Vote for the feminine party**	Dummy variable. Coded 1 if the judge votes for a party where a mother is represented. A vote for a party representing the father alone or the public interest is coded 0.
Individual level explanatory variables		
Gender	**Female judge**	Dummy variable: Coded 1 if the judge is a woman and 0 if the judge is a man.
Critical actors	**Critical actors**	One dummy variable for each relevant judge. Coded 1 if the voting judge is the judge in question. All other coded 0.
Panel level explanatory variables		
Gender and critical mass	**At least one woman present; Female majority in panel; Presiding judge is female**	Dummy variables. Coded 1 if the panel consists of a majority of female judges; one woman is present; or the presiding judge in the panel is female. All other coded 0.
	Percentage of women in panel	Metric variable. Percentage of women in each panel. Values: 0, 0.2, 0.4, 0.6, 0.8 and 1.
Critical actors' peer effect	**Critical actors present in the panel**	Dummy variable. Coded 1 if one of the identified "critical actors" are present. All other coded 0.
Case particular explanatory variables		
Case particularities	**Feminine party's lawyer is female; Opposing party's lawyer is female**	Dummy variables. Coded 1 if the lawyer of the feminine party is female. All other coded 0.
	Opposing party represents the public interest (child welfare service)	Dummy variable. Coded 1 if the party opposing the female party is represented by the public. In all other cases, the opposing party is the child father. These are coded 0.
Time	**Years since 1968**	Metric variable. Years since the appointment of the first female judge in *Høgsterett* in 1968.
	After 1987 (dummy)	Dummy variable. Coded 1 if the case was conducted after 1987 and 0 if it was coded before 1978.

Dependent variable

As can be seen in Table 4.1, the dependent variable in this analysis is a dummy variable measuring whether an individual judge has voted in favour of the claim presented by a feminine party or the opposing party in cases of child custody.

A "feminine party" is here understood as any party where *a woman* (i.e. a mother) is represented[50]. Specifically, the concept includes observations where a) a mother stands alone, and b) where both the mother and the father is represented together. On the other hand, the concept excludes any party where no mother is present, such as where c) only the father is represented, or d) the public interest (the Child Welfare Service) is represented.

Essential to the coding of this variable lies a qualitative evaluation within each case of which party the judges in the judicial panel has ruled *in favour of*. At the start of every trial, both parties enter a formal claim. The evaluation consists of comparing each party's initial claim to the verdict that the judge voted in favour of. Although the verdict is seldom equal to any of the two initial claims, it is always closer to one of the two. That party, whose initial claim is nearest to the verdict that the individual judge voted for, is considered to have been voted in favour of[51].

Explanatory variables

The main explanatory variable on the individual level is a dichotomous measure of the voting judge's gender.

To measure the collegial effects of gender, four measures of the panel gender balance is included in the analysis. The first three of these are dummy variable. The first identifies all panels where there is at least one woman present. The second identifies panels with female majority and the third identifies panels where the presiding judge is female. Finally, a metric variable is included to measure the percentage of women in the panel

Critical actors

As described in the analytical strategy presented in the previous chapter, Critical actors are identified as part of the analysis. To do this an individual level dummy variable is coded for all judges who have partaken in ten or more cases within this selection. To measure any collegial effects of the critical actors, panel level dummies are coded to identify the panels in which the relevant critical actor is present.

[50] Duly noted, in absolutely all cases the child's mother was represented on one or the other side.

[51] This is not unproblematic. According to *Høgsterett*, its purpose is "to ensure uniformity, clarity and development of the law". As such, arguably, judges in *Høgsterett* are supposedly more concerned with precedent than the actual case at hand.

Case particularities

Dummy variables have been coded to identify panels in which either the feminine party or the opposing party is represented by a female lawyer.

To measure time, a metric variable has been coded, centred at the first year of the selection, which is the year when Lily Bølviken was appointed the first female judge in *Høgsterett*. In effect, cases set before the court in 1968 are given the value 0 on this variable, 1969 has got the value 1, etc. until 2011, which is represented by the value 43. Centring the variable to a zero-point that can be observed in the selection is important for meaningful interpretation of the coefficients in the analysis.

An additional dummy variable has been coded to provide a crude measure of the most and least recent time-period of the selection. The dummy "After 1987" roughly splits the selection in two parts over the temporal dimension.

4.4.2 Control variables

Individual level control variables

As can be seen in Table 4.2 on the following page, Leftist appointment is a measure of whether the justice was appointed by a government under a Labour party prime minister.

Grendstad et al. (2011b) found ideology, operationalised as a function of whether the justice had been appointed by a leftist government to be a significant factor in explaining the outcome of a selection of 63 non-unanimous cases of economic disputes where one of the parties represented the *public interest* in *Høgsterett* in the period between 2000 and 2007. Judges appointed by leftist governments were more likely to vote for the public party than others were. Similarly, Bentsen (2012) found that the propensity of justices in Høgsterett to dissent rises with *the ideological spread* in the panel in which they sit. Although it is hard to theorise an effect of the latter factor on the final outcome of cases of child custody, it is an indication of the importance of including a measure of judges' ideology in analyses of judicial behaviour within the context of *Høgsterett*.

Notably, as can be seen in Table 4.2, the variable has been given an operational definition including all governments with a Labour party prime minister. This means that the present

ruling coalition of Labour, the Socialist Left Party and the Centre Party is included although the Centre Party is not traditionally understood to be a left wing party.

Table 4.2: Operationalisation of all control variables

Theoretical concept	Variable	Operationalisation
Individual level control variables		
Ideological background	**Leftist appointment**	Dummy variable: Coded 1 if the judge was appointed by any government under a Labour party prime minister. All other coded 0.
Geographical background	**Oslo-born**	Dummy variable: Coded 1 if the judge was born in Oslo. All other coded 0.
Experience	**Age (c)**	Metric variable: The original values were a measure of each judge's age at the time of voting. The variable has been centred at its mean value of ~60 (c).
	Seniority	Metric variable. Counts the number of years the judge has been serving *Høgsterett* at the time of vote.
	Year of graduation (c)	Metric variable. The original values were the years of graduation for each individual judge. The variable has been centred at its mean value of ~1952 (c).
	Graduated after 1968	Dummy variable. Coded 1 if the judge graduated after the year of employment for the first female justice in *Høgsterett*. All other coded 0.
Former work place	**Previously employed as judge; by the Attorney General of Civil Affairs; by the Director of Public Prosecution; or by the Legislation department**	Dummy variables: Coded 1 if the judge previously held the position in question. All other coded 0.
Panel level control variables		
Geographical background	**Oslo-majority**	Aggregate from individual level. Dummy variable: Coded 1 if three or more of the judges in panel are Oslo-born. All other coded 0.
Experience	**Mean graduation year in panel (c)**	Aggregate from individual level. Metric variable.
Case particularities	**Consensus**	Dummy variable. Coded 1 if there are no dissenting votes in the panel. Coded 0 if one or more judges dissent.

The variable Oslo-Born is a dummy, identifying judges who are born in the capital of Norway, Oslo. In the aforementioned analysis, Grendstad et al. (2011b) also found that the justice's geographical background, with the operational definition of whether or not the individual justice was born in Oslo had a significant effect on the outcome of the cases in their selection. Specifically, Oslo-born judges were more likely to vote for the public interest.

Age is a metric variable, identifying the age of the judge at the time of voting. Studying criminal cases before Høgsterett in the period from 1996 to 2011, Bentsen (2012) found age to have a significant and negative effect on the Supreme Court judges' propensity to take out dissent. As study of child custody cases conducted by Drammens tidende (2009) on three local district courts also indicated that age had an effect on judges' votes. This variable has been centred at its mean value of ~60 to enable meaningful interpretation of its coefficients in the analysis. This means that a value 1 on the variable is equal to one year older than the mean age, and consequently the values span both above and below 0.

Seniority is a metric variable measuring how many years the judge has served *Høgsterett* at the time of voting. Skiple (2012) found judicial experience to significantly affect judges propensity to vote in favour of the public party. In this thesis, seniority has been added to control for potential similar effects. Because this variable has a natural zero-point that is observable within the selection, there is no need to centre this variable.

Year of graduation[52] is a metric variable measuring the year a judge was graduated. The farther back in time a judge has graduated, the longer time has this judge been acting her or his profession knowing of a female judge in the Supreme Court. This variable has been centred at its mean value of ~1952 to enable meaningful interpretation of its coefficients in the analysis. This means that a value 1 on the variable is equal to graduation one year later than the mean graduation year, and consequently the values span both above and below 0.

The variable "Graduated after 1968" is a dummy variable which provides a rougher measure of the aforementioned effect.

The four Former work place variables are dummies identifying votes cast by judges who have had the relevant work experience. There are good reasons to assume that the judges' former work place might affect their perspectives and thus their voting pattern. Jacobsen (2012) found that judges who had been formerly employed in the Legislation Department in the Ministry of Justice and Public Security were significantly more prone to vote in favour of the

[52] This variable and the other measures of age and seniority has been tested for multicollinearity as accounted for in section 3.2.7

state in both unanimous and non-unanimous civil cases where the state was a party in the period from 1990 to 2010. This was also the case for judges who had been formerly employed as judges. Skiple (2012) found similar effects for the legislation department, but not for former employment as judge. Bentsen (2012) found former employment by the Attorney General of Civil Affairs[53] to significantly affect judges' propensity to dissent. Similarly, former employment by the Director of Public Prosecution[54] had a marginally significant effect on the judges' propensity to dissent.

On the panel level, the "Oslo-Majority" variable and the "Mean graduation year in panel" variables are aggregates of aforementioned individual level variables. The first identifies panels in which more than 50% of the judges are Oslo-born. The latter is the mean graduation year within each panel. These variables are constants within each panel and vary only between panels.

Finally, Consensus is a dummy identifying panels in which all judges voted the same.

[53] *Regjeringsadvokaten*
[54] *Riksadvokaten*

5 Empirical analysis

5.1 Descriptive statistics

Most of the variables included in the analysis are dichotomous, coded as dummy variables, ranging from 0 to 1. Thus, the variables function as "on/off switches" for the relevant phenomena (Midtbø, 2007:33-34). Because of this property, the mean value of a dummy variables can be interpreted as the proportion of observations in the sample that is equal to 1 (Steele, 2009a:35). The only non-dichotomous variables are metric measures of time in one way or another.

5.1.1 Dependent and explanatory variables

Table 5.1: Descriptive statistics of dependent and explanatory variables[55]

			Female		Male		Total	
	Min	**Max**	**Mean**	**S.d.**	**Mean**	**S.d.**	**Mean**	**S.d.**
Dependent								
Vote for feminine party	0	1	.41	.49	.46	.50	.45	.50
Individual level								
Female judge	0	1	1	0	0	0	.21	.41
Panel level								
Collegial factors								
One woman present	0	1	1	0	.60	.49	.68	.47
Female majority	0	1	.22	.42	.03	.16	.07	.25
Presiding judge is female	0	1	.37	.48	.17	.37	.21	.41
Percentage of women in panel	0	1	.39	.20	.16	.16	.21	.19
Case particularities								
Feminine party's lawyer is female	0	1	.23	.42	.14	.35	.16	.37
Opposing party's lawyer is female	0	1	.21	.41	.17	.38	.18	.38
Opposing party represents the public interest	0	1	.35	.48	.36	.48	.35	.48
Years since 1968	0	43	~24	~11	~18	~11	~19	10.98
After 1987	0	1	.64	.48	.46	.50	.49	.50

Note: There are in total N*=750 units at the individual level, 155 for female part and 595 for the male part of the sample. At the panel level, there are N=150 units, 31 for the female part and 119. There is no loss of observations. Source: Doranoh (Grendstad et al., 2013a)*

[55] Descriptive statistics for variables identifying critical actors are presented where these have been identified. See section 5.2.3 *Step 3: Identifying critical actors*

Dependent variable

The dependent variable in this analysis is dichotomous, coded as a dummy variable. We see from Table 5.1 that 45% of all votes cast in this sample have been cast in favour of the feminine party. There is a slight difference between the mean vote of male and female judges, where only 41% the female judges' votes have been cast in favour of the female party.

The slight overall tendency for judges to vote in disfavour of the female party in this selection immediately seems to counter the literature in the field[56]. There, the tendency has been a slight favouring of the female party as opposed to a male party. However, as we know from the data section, the present selection includes both cases where the opposing party is a representative of the public interest, and cases where the opposing party is the child father. This is the reason for the apparent divergence.

Figure 5.1: Mean vote split by opposing party

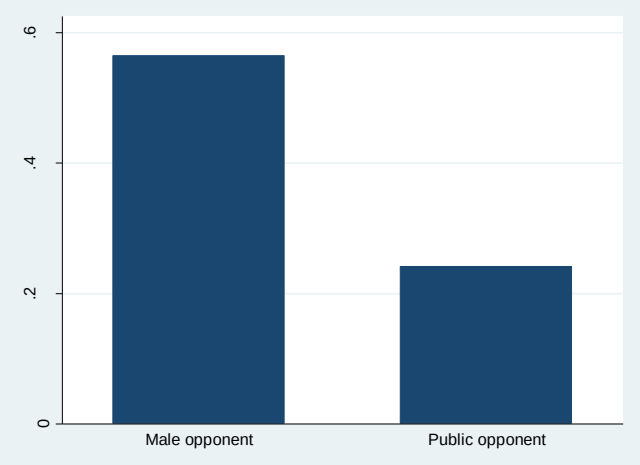


Figure 5.1 shows that there is a large difference between the mean vote[57] in cases where the opposing party represents an individual man and cases where the opposing party represents the public interest. Specifically, where the opponent is an individual man, 57% of the votes are cast in favour of the female party, while only 24% of the votes are cast in favour of the female party where the opponent is a representative of the public. This tendency is more similar to what has been found in similar types of cases in the lower courts, where there also is a slight tendency for the female party to be favoured[58].

Individual level

Looking back at Table 5.1, on the individual level, only 21 % of the total number of votes is cast by female judges. Table 5.2 below shows why this tendency should not be surprising. The first female judge in *Høgsterett* was justice Bølviken in 1968 – which is the very year marking the start of the case selection for this thesis. She was accompanied by justice

[56] See e.g. Drammens tidende (2009) and Skjørten (2005) in Ministry of Children and Equality (2008:22). These present calculations for the lower Norwegian courts, not *Høgsterett*.
[57] Duly noted, this is the total vote (both female and male judges).
[58] See previous footnote.

Schweigaard Selmer in 1971 and justice Holmøy in 1976. However, although Bølviken left her position in 1984, a fourth female judge was not appointed until justice Gjølstad entered the collegium in 1988. Thus, for the first twenty-year period since 1968, no more than three female judges had served the court at the same time, and not until justice Gussgard was appointed in 1990 did the number of female judges increase to more than three.

Table 5.2: All permanent female judges in *Høgsterett,* by year of commencement

Start date	Name	Birth	Birth city	Ending date
31.05.1968	Lilly Bølviken	1914	Arendal	30.03.1984
01.01.1971	Elisabeth Schweigaard Selmer	1923	Christiania*	30.09.1990
01.10.1976	Vera Louise Holmøy	1931	Oslo	30.04.2001
01.09.1988	Liv Gjølstad	1945	Tønsberg	2015 at the latest
01.10.1990	Karenanne Gussgard	1940	Sandefjord	01.02.2010
15.08.1994	Kirsti Coward	1940	Kristiansand	31.12.2010
06.10.1997	Karin Maria Bruzelius	1941	Lund, Sweden	28.02.2011
01.05.2000	Nina Frisak	1950	Oslo	14.11.2001
01.05.2001	Ingse Stabel	1946	Oslo	2016 at the latest
01.08.2004	Toril Marie Øie	1960	Oslo	2030 at the latest
01.04.2007	Hilde Indreberg	1957	Oslo	2027 at the latest
15.08.2009	Bergljot Webster	1966	Oslo	2036 at the latest
09.08.2010	Kristin Normann	1954	Oslo	2024 at the latest
01.10.2010	Ragnhild Noer	1959	Oslo	2029 at the latest

*Note: Except justice Noer, all female judges have cast one or more votes in the selection. The judges marked by grey background have cast votes in ten or more of the cases in the present selection of cases. Temporary assignments are not included. *Christiania was the name of the capital, Oslo until 1925.*

Collegial factors

In the twenty years following 1990, the number of women appointed to the collegium has increased considerably. Looking back at Table 5.1, we see that there has been at least one female judge present in 68% of the panels in this selection, with a mean panel-percentage of female judges of 21%. Female majority is seen in only 7% of the cases. In 21% of the cases, the presiding judge was female. Naturally, these numbers are higher for the votes cast by female judges, where – by definition – at least one female judge has to be present in all observations.

Case particularities

Further, Table 5.1 shows that the feminine party's and the opposing party's lawyer was female in 16% and 18% of the cases respectively. These numbers are slightly higher for votes cast by female judges. This is likely to be due to the time factor. Although the number of female Supreme Court lawyers is still very low (Aschehough, 2011), during the same period that more women have entered the judicial collegium, there has also been a slight increase in female Supreme Court lawyers. Thus, the likelihood of a female lawyer present increases with

the presence of female judges. Additionally, 35% of the votes are cast in cases where the opposing party is a representative of the public, while in the remaining 65% the opponent is the child father. This is equally distributed for both female and male votes.

The last two explanatory variables are measures of time, assuming that the perception of gender roles will have changed over time. The first of these, *Years since 1968*, has the minimum value 0 (for 1968, the year of appointment for the first female *Høgsterett* justice and first year of the selection) and the maximum value 43 (for the last year in the selection, 2011). The total mean vote in this selection was cast 19 years after 1968, i.e. in 1987. The standard deviation for this measure is approximately 11, indicating that the cases are evenly spread over the period. Note that there is a slight difference between the mean male and female vote on this variable. This is due to more female judges being employed in the court in the latter than the earlier years of the period.

The last variable shows that 49% per cent of the cases are dated to a year between (and including) 1988 and 2011. The remaining 51% are dated to a year between (and including) 1968 and 1987. The variable thus splits the observations in two approximately equally sized groups over the time scale, as intended. For the female votes, however, as many as 64% are cast in the latter period.

5.1.2 Control variables

Table 5.3: Descriptive statistics of control variables

			Female		Male		Total	
	Min	**Max**	**Mean**	**S.d.**	**Mean**	**S.d.**	**Mean**	**S.d.**
Individual level								
Leftist appointed judge	0	1	.60	.49	.66	.48	.64	.48
Oslo-born judge	0	1	.43	.50	.38	.49	.39	.49
Age (c)	-19.68	10.32	-2.71	6.75	.71	6.49	~0	6.69
Seniority	0	28	8.12	5.96	8.87	6.63	8.71	6.50
Year of graduation (c)	-32.40	39.60	6.91	10.77	-1.80	13.92	~0	13.78
Graduated after 1968	0	1	.27	.44	.12	.32	.15	.36
Previously employed as judge	0	1	.50	.50	.44	.50	.45	.50
Previously employed by the Attorney General of Civil Affairs	0	1	.05	.21	.23	.42	.19	.39
Previously employed by the Director of Public Prosecution	0	1	0	0	.08	.27	.06	.24
Previously employed by the Legislation department	0	1	.57	.50	.34	.47	.38	.49
Panel level								
Collegial factors								
Oslo-majority	0	1	.34	.48	.32	.47	.33	.47
Mean graduation year in panel (c)	-26.40	27	4.86	11.71	-1.27	11.96	~0	12.16
Case particularities								
Consensus	0	1	.76	.43	.74	.44	.74	.44

Note: There are in total N=750 units at the individual level, 155 for female part and 595 for the male part of the sample. At the panel level, there are N=150 units, 31 for the female part and 119. There is no loss of observations. Source: Grendstad et al. (2013a)

Studying judges' background in the present sample, Table 5.3 shows that 64 % of the total number of votes is cast by judges appointed by leftist-oriented governments. The slightly higher percentage for male votes on this variable is likely due to the historical fact that there were more leftist-oriented governments in Norway in the early period than in the latter part of the period. Additionally, 39% of the votes are cast by Oslo-born judges, a percentage that is somewhat higher (five percentage points) for the female votes than for the male.

Because the age variable is centred, its mean value is zero. The total standard deviation tells us that the mean spread around the mean age is approximately 6.5 years. Not surprisingly, we can also observe that the mean female vote is cast by a woman who is 2.71 years younger than the total mean, and almost 3.5 years younger than for the mean vote cast male judges. For the seniority variable, the difference between the two gender groups is more even, with less than a year difference between the two groups.

A bit more surprising is to see that the difference in graduation year is as high as it is. The mean female vote is cast by a judge who has graduated almost seven years later than the total mean, almost nine years after the mean graduation for male judges. This could indicate that the female judges in this panel have spent quite a lot less time between their graduation and their accentuation into the collegium of the Supreme Court, than their male colleagues have. We should remember, however, that this is the mean of the 750 votes. It is clearly biased with the vote of some judges observed more than thirty times, while many of the judges have cast less than 10 votes in the selection.

While only 15% of the votes are cast by judges who graduated after 1968, the remaining 85% are cast by judges who had never heard of a female Supreme Court judge before their graduation year. Clearly more female votes are cast by judges graduated after 1968.

While half of the female votes are cast by judges previously employed as judges, only five per cent are cast by judges previously employed by the Attorney General. While the first of these does not differ very much from the male mean, the latter is clearly much lower. This is also the case for the Director of Public Prosecution, where the judges behind 8% of the male votes have been employed, while the same number for female votes is zero. For employment at the Legislation department, the tables are clearly turned.

33% of all votes are cast in panels where the majority is Oslo-born. Not surprisingly, this is similar for the two gender groups. Notably is, however, that the female votes are cast in panels where the mean panel graduation year is notably lower than the similar for male votes. This is, however, explainable by the big difference at the individual level.

Lastly, 74% of the votes are cast in unanimous panels. This will be further elaborated upon in the coming section.

5.2 Multi-level analysis

5.2.1 Step 1: Empty models and intra-class correlations

Table 5.4: Four empty two-level models of judges' propensity to vote for the feminine party

	Full data set						Only cases of dissent					
	Empty model 1			**Empty model 2**			**Empty model 3**			**Empty model 4**		
	Intra-panel dependencies			Intra-individual dependencies			Intra-panel dependencies			Intra-individual dependencies		
	N = 150 (panels) N = 750 (votes)			*N* = 88 (judges) N = 750 (votes)			*N* = 39 (panels) N = 195 (votes)			*N* = 61 (judges) N = 195 (votes)		
	coeff.	s.e.	p	coeff.	s.e.	p	coeff.	s.e.	p	coeff.	s.e.	p
intercept	- .81	.58	.165	-.18	.10	.060	.01	.15	.943	.05	.20	.820
Random												
level 2 res. variance (σ^2_{uo})	36.48			.17			.02			.71		
ICC	.92			.05			.01			.18		
LR test vs. flat logistic regression	*** p<.001			*** p=.006			.460			** .011		
AIC / BIC	588.643 / 597.883			1030.224 / 1039.464			274.312 / 280.858			269.008 / 275.554		

Note: The dependent variable is individual judges' propensity to vote for the feminine party. The analysis is run using the xtmelogit command with 24 integration points in Stata.
**** p<.01, ** p<.05, * p<.10.*

The first step in the analysis is to estimate the *intra-class correlation (ICC)* in an empty model assuming votes to be grouped within the panels in which they are cast. This can be seen in *Empty model 1* in Table 5.4. As explained in section 3.3.2, the ICC in this estimation is a statistical measure of how much of the variance that happens *between* the five-judge panels, rather than *within* the panels. A naïve interpretation of the ICC presented in the table is that 92% of the variance in the dependent variable needs to be explained by factors on the second level of analysis (related either to the composition of the panel or to particularities of the individual case at hand). This leaves 8% per cent of the variance, remaining to be explained by factors related to the individual judges.

This result is well within the limits of what Acock (2010:239) describes as an "extremely high level of agreement", and needs a bit of explanation. *First,* consensus is the norm of the court. Cases of dissent make out approximately 20% of all cases in general (Aschehough, 2012). In the present selection, this number is somewhat higher with dissenting votes in one out of four cases[59]. *Second,* even when there is dissent in a panel, normally only one of the judges

[59] To be precise, there is one or more dissenting votes registered in 41 out of 150 (26%) in this selection.

dissent. This can be illustrated by comparing the percentage of individual dissents with the previously calculated percentage of dissent cases: Although there is dissent in 26% of the cases, only 8%[60] of the votes are actual dissents. The remaining 92% vote with the majority of the panel, which is the same estimation as was given by the ICC.

In other words, the ICC simply tells us how many votes are cast in consensus with the majority fraction of the class we define. This can help understanding and interpreting the ICC. In this particular case, the class was defined by the panels, but in *Empty model 2* in Table 5.4, analysed in the following section, the class is defined by the judges rather than the panels as a measure of whether the judges "agree" with themselves.

Although the ICC does not tell us anything we could not have found by simply calculating the rate of dissent, it is still a useful statistic. The ICC is a statistic commonly used to evaluate whether or not we should assume a certain hierarchical structure in the analysis of our data. An ICC close to 92% tells us that the data does have a hierarchical structure, with parts of the explanation belonging to two different levels. This is a strong statistical argument in favour of the chosen multilevel structure (see Hox (2010:4-5)), rather than a "flat" logistical analysis. This conclusion is further supported by a likelihood-ratio test showing that the explanatory power of the hierarchical model is significantly improved from that of a "flat" logistic model[61].

Intra-individual dependencies in the full data set

A similar empty model with the data clustered in judges rather than panels gives a quite different and interesting result. The intra-class correlation is calculated to 0.05, naïvely interpreted to meaning that five per cent of the variance in the dependent variable is due to intra-individual dependencies. In the previous section, we calculated an expected 8% of the variance to be due to changes related to neither panel composition nor case particularity. The present result is only two percentage points lower and ought to be seen as pretty close. Notably, since we assume two completely different data structures in the two calculations, it would be unreasonable to expect the two calculations to overlap completely.

[60] The frequency of individual dissents is 61 out of 750 votes (8%).
[61] This is explained by Hox (2010:47-50)

By comparing the information criteria AIC and BIC[62] for the first two empty models in Table 5.4, we see that the goodness of fit is clearly lower when we group the votes by judges rather than in panels. This is a strong statistical argument for preferring the chosen two-level design where votes are grouped by panels rather than judges.

Auto-dependencies when including only cases of dissent

In the two last columns in Table 5.4, we see similar empty models calculated when only cases of dissent are included. In these models, we should note that level 1 N is now reduced from 750 to only 195. On the two potential second levels, the number of panels is reduced to only 39 from 150, and the number of judges is 66, rather than 88.

Interestingly, the ICC in this is calculated to only 1 % when the votes are clustered in panels. This is a striking difference from the 92% intra-class correlation in the full data set including consensus cases. Not surprising then, the LR-test also shows that the multi-level structure does not provide any significantly improved explanatory power. Following this trail, in the fourth model in Table 5.4 the votes in *only cases of dissent* are grouped by judges rather than by panels. The ICC is now calculated to 18%, indicating a significant variance explained at the individual level in this part of the data set, and the LR-test shows that the explanatory power is significantly improved with this multi-level structure rather than a flat analysis.

Within the selection of dissent cases, the AIC/BIC is slightly improved in the model where votes are grouped by judges, opposite from that of the full data set, although the differences are notably smaller. Comparing the two selections, however, we see that the models fit much better when including only cases of dissent.

Summing up: a substantial interpretation

The substantial interpretation of these numbers is that – unsurprisingly – in cases of dissent, factors related to the judges themselves (as opposed to the particularities of the individual cases or the composition of panels) are much more prominent than in cases of consensus. Factors related to individual judges' background, such as their gender, where they grew up or former work places should be expected to explain twice as much of judges voting pattern when a judicial panel is split by dissent. See i.e. Bentsen (2012) for an analysis of why judges take out dissent in *Høgsterett*.

[62] See 3.3.2 (Step 1 in the analytical strategy) for a discussion of AIC/BIC.

5.2.2 Step 2: The general effect of gender

Table 5.5: Exploring the general effect of gender on the individual level

Variables	**Block 1 Gender**			**Block 2 Individual ideology**			**Block 3 Geographical background**			**Block 4 Experience**			**Block 5 Former work place**[63]		
	coeff.	s.e.	p	coeff.	s.e.	p	coeff.	s.e.	p	coeff.	s.e.	p	coeff	s.e.	p
Intercept	-.64	.60	.28	-.78	.64	.23	-1.00	.62	.11	-.82	.70	.24	-.51	.69	.456
Female judge	-.86**	.42	.04	-.83*	.42	.05	-.92**	.42	.03	-.80*	.45	.08	-.79*	.47	.096
Leftist appointed judge				.19	.33	.56									
Oslo-born judge							.89***	.34	.01	.81**	.35	.02	.93***	.36	.009
Age of judge (c)										-.05	.06	.47			
Seniority of judge										-.03	.04	.41			
Year of graduation (c)										-.09*	.05	.07	-.04	.02	.115
Graduated after 1968										-.01	.76	.99			
Prev. emp. as judge													-.68**	.36	.058
Prev. emp. by the Attorney General of Civil Affairs													-.46	.54	.389
Prev. emp. by the Director of Public Prosecution													-.12	.73	.868
Prev. emp. by the Legislation department													-.46	.37	.219
Random effects															
Level 2 residual variance	38.03			38.11			39.31			42.30			41.03		
Intra-class correlation	.920			.921			.923			.928			.926		
LR-test	.04**			.56			>.01***			.21			.284		
AIC *BIC*	586.1908		600.051	587.8422		606.3225	580.9869		599.4671	583.0472		620.0078	582.0387		623.6194

*Note: The dependent variable is individual judges' propensity to vote for the feminine party. There are N=750 observed votes within n=150 panels with five votes in each. The analyses are run using the xtmelogit command with 24 integration points in Stata. *** p<.01, ** p<.05, * p<.10. "(c)" indicates variable centred at mean. LR-tests compare with the previous block which significantly contributed to the explanatory power of the model. In addition, where the entire block itself does not contribute to the explanatory power of the model, but one or some of the explanatory variables do and are included in the following blocks, the LR-test compares with downscaled versions of the block, because comparison to a rejected model is meaningless. Stata output for the downscaled versions can be found in the appendix for reference.*

[63] The Legislation department is *Lovavdelinga*, The Attorney General of Civil Affairs is *Regjeringsadvokaten* and The Director of Public Prosecution is *Riksadvokaten*.

Gender on the individual level

A different voice: Male judges are more likely to vote in favour of the feminine party

The first block of Table 5.5 shows an estimate of the likelihood for female and male judges to vote for the feminine party in cases of child custody when controlling only for the group structure chosen in the first step of analysis. The tendency is negative and significant on a five per cent level, both in terms of a z-test and an LR-test comparing this block to the previously presented empty model. This indicates that female judges are less likely to vote in favour of the feminine party than male judges are. Specifically, the odds of a female judge voting in favour of the female party in cases of child custody are 58% lower than that of a male judge[64].

As expected, the intra-class correlation (ICC) remains approximately unchanged from the empty model to block 1 in Table 5.5. Because the ICC mainly relies on differences in the level-2 variance term in logistic analyses, this implies that the gender distribution does not vary a great deal from panel to panel (see section 3.3.2 and Steele (2008:18-21) for reference).

When controlling only for the group structure of the data, we find strong support for hypothesis 1, that *"Male and female judges vote significantly different from each other"*, thus the results so far favours all other approaches than the Organisational one (see Table 2.2, summarising the analysis' expected effects).

Despite the significant gender effect, however, we do not find any support for hypothesis 2, that *"Female judges are significantly more likely to vote in favour of the female party*". The effect goes in the opposite direction, thus disfavouring the assumption made on behalf of the Representational approach. This will be further discussed later on.

Hypothesis 3, which in effect is the null-hypothesis of the two previously discussed hypotheses, assuming that *"Any difference found in the voting pattern of male and female judges is purely random"* does not find any support so far.

[64] The exponential regression coefficient in logistic regression analysis can be interpreted as an odds ratio (Rabe-Hesketh and Skrondal, 2008:252-3, Leckie, 2010b:P6.6.2). The estimated percentage decrease in odds can thus be calculated by the formula "$100 \times (Exp(\beta) - 1)$", where β is the coefficient (Rabe-Hesketh and Skrondal, 2008:249, Eikemo and Clausen, 2007:92, Leckie, 2010b:P6.6.2).

Judges born in the capital are more likely to vote in favour of the feminine party

In the following four blocks of Table 5.5, this "gender effect" is tested by controlling for factors related to each individual judge. In every block, a new group of individual-level control variables is introduced. In block 2 we find that the chosen measure of *ideology*, operationalized as whether the judge was appointed by a leftist oriented government, does not provide any significant prediction of judges' voting pattern in cases of child custody.

However, in the third model, the measure of *geographical background*, operationalized as whether or not the judge is born in Oslo has a clearly significant effect on judges' voting pattern. This block tells us that, controlling for the effects of gender and the chosen group structure, the predicted odds of an Oslo-born judge to vote in favour of the feminine party is as much as 144% higher than a judge born elsewhere.

Both the LR-test[65] and the two information criteria supports the conclusion that the geography variable improves the model and should be included in the further analysis. The variable does not, however, seem to have any significant effect on the estimated coefficient for the gender variable.

Block 4 and 5 are both abandoned as a whole based on both the LR-test and a comparison of the two information criteria against block 3. In block 4[66], controlling for judges' experience, however, we see that the judges' year of graduation provides a marginally significant effect on the outcome. A new analysis of this block was therefore run, letting only *year of graduation* measure judges' experience (see appendix, section 7.4). In this analysis, although the variable's coefficient is still only marginally significant (p=.052), the LR-test showed a significant improvement of explanatory power. Based on the LR-test, the graduation year variable is therefore included in further analysis, although the two information criteria indicate that it has little effect on the model's goodness-of-fit.

[65] As noted in the table (Table 5.5), the test is run vs. the "previous significant model", understood as the previous model with significantly improved explanatory power. In this case, block 1.

[66] Particularly with regard to block 4, in which several of the explanatory variables can seem to measure the same phenomenon, at all steps of the process, the data has been tested for *multicollinearity*, as accounted for in section 3.2.7.

As a similar case, in block 5, controlling for former work experience, the variable measuring whether the judges have been previously employed as judges in the lower courts before entering the high court seems to have a significant effect on the outcome. This block was also tested again, with only the significant variable (see appendix, section 7.5). In this reduced block, neither the coefficient, nor the LR-test came out significant and the information criteria indicated no significant change in the model's goodness-of-fit. This variable was thus rejected together with the rest of block 5.

Summing up: Gender and geography affects judges' voting pattern

On the individual level, we are thus left with two significant predictors: *Gender*, measured as being female and *geography*, measured as being born in the capital. These two variables are included in the further analysis, exploring the effect of gender on the collegial level.

Having controlled for ideology, geographical background, experience and former work place, we see in Table 5.6 that female judges are significantly less likely to vote in favour of the female party than are male judges. Similarly, judges born in Oslo, as opposed to anywhere else[67], are more likely to vote in favour of the female party. The analysis also indicates that the more recently a judge graduated, the lower are the odds that he or she votes in favour of the female party. This last effect is only significant on a ten per cent level, and we should therefore be careful with assuming that we can generalise based on this result.

Table 5.6: The effect of gender, not controlling for collegial/panel effects or critical actors

	coefficient	standard error	p-value	Summary
intercept	-1.03	.63	.101	Level 1 units (*N*) 750 (judges)
Female judge	-.82*	.43	.057	
Born in Oslo	.84**	.34	.014	Level 2 units (N) 150 (panels)
Year of graduation (c)	-.04*	.02	.052	
level 2 residual variance		40.17		LR test vs. flat logistic regression: p<.01
intra class correlation		.924		
LR-test vs. empty model		<.01		
AIC / BIC		579.0732 / 602.1735		

[67] Notably, justice Bruzelius is even born in another country.

*Note: The dependent variable is individual judges' propensity to vote for the feminine party. There are N=750 observed votes within n=150 panels with five votes in each. The analyses are run using the xtmelogit command with 24 integration points in Stata. *** p<.01, ** p<.05, * p<.10. "(c)" indicates variable centred on mean.*

Duly noted, the intra-class correlation calculated in the empty model gives a clear indication that whatever effects we find on the individual level can contribute to explain only as much 8% of the total variance in the outcome variable. As expected[68], this has not changed significantly although we have found three explanatory variables on the individual level to significantly affect the outcome in the dependent variable. The main remainder of explanatory power is left on the collegial or panel level, which is the next step of the analysis.

So far, the analysis of individual level factors supports hypothesis 1 and rejects hypotheses 2 and 4. Thus, the results weaken the organisational approach and also contradict the direction of the gender effect as it is expected on behalf of the representational approach.

[68] The level-1 residual variance is fixed because of the logistical recoding of the dependent variable. See section 3.3.2.

Gender on the collegial level

Table 5.7 Exploring the effect of gender on the collegial level

Variables	**Block 1 Gender**			**Block 2 Aggregate controls**			**Block 3 Case particularities**			**Block 4 *Time***		
	coeff.	s.e.	p	coeff.	s.e.	p	coeff.	s.e.	p	coeff.	s.e.	p
Intercept	-1.95*	1.07	.070	-10.39	106.59	.922	.86	1.04	.406	-.76	1.19	.521
Female judge	-.86*	.44	.050	-.83*	.43	.056	-.86**	.43	.043	-.95	.42	.025
Born in Oslo	.83**	.34	.015	.79**	.34	.021	.85**	.34	.012	.89	.34	.008
Year of graduation (c)	-.04*	.02	.050	-.04*	.02	.090	-.02	-.02	.319			
At least one woman pres.	-.09	2.33	.968									
Female majority	-5.27	4.06	.194									
Percentage of women	6.44	7.50	.391									
Presiding judge female	.01	1.60	.990									
Majority born in Oslo				2.08	1.32	.115						
Panel mean grad. year				.01	.06	.935						
Feminine party has female lawyer							.75	1.55	.630			
Opposing party has female lawyer							-.43	1.53	.780			
Opposing party is public							-5.15***	1.35	.000	-5.89***	1.45	.000
Consensus							-.28	1.18	.814			
Years since 1968										.15	.10	.114
Last half of the selection										-2.27	2.06	.271
Random effects												
Level 2 residual variance	38.72			39.70			32.88			30.84		
Intra-class correlation	.923			.924			.909			.904		
LR-test	.562			.268			.001***			.258		
AIC *BIC*	584.095		625.6756	580.4362		612.7767	566.6378		603.5984	563.2667		595.6072

*Note: The dependent variable is individual judges' propensity to vote for the feminine party. There are N=750 observed votes within n=150 panels with five votes in each. The analyses are run using the xtmelogit command with 24 integration points in Stata. *** p<.01, ** p<.05, * p<.10. "(c)" indicates variable centred at mean. LR-tests compare with the previous block which significantly contributed to the explanatory power of the model. In addition, where the entire block itself does not contribute to the explanatory power of the model, but one or some of the explanatory variables do and are included in the following blocks, the LR-test compares with downscaled versions of the block, because comparison to a rejected model is meaningless. Stata output for the downscaled versions can be found in the appendix for reference.*

Judges are not affected by the gender of their peers

The interpretation of Table 5.7 can be conducted in a similar fashion to that of Table 5.5. The main difference is that, in addition to the inclusion of the three significant variables from the individual level analysis, the variables entered in this part of the analysis are positioned on level 2 of the analysis. The observations on these variables vary only *between* panels, and are constant *within* the panels.

Already in block 1, we see that although they are seemingly affected by their own gender, *Høgsterett* judges are not affected by the gender of their peers in cases of child custody. This block introduces four measures of the gender composition of panels, none of which prove significant[69]. This leads to the dismissal of hypotheses seven through ten, as presented in Table 2.2 on page 14. There are no empirical grounds to assume that the gender effect is affected by the presence of a critical mass of female judges (H7); that the gender balance in a panel affects the voting pattern of its judges when controlling for individual gender (H8); that the voting pattern of all judges in a panel is affected by having just one female judge present (H9), nor; that the gender of the presiding judge affects the voting pattern of judges in the panel (H10).

The results in the first block are, however, in accordance with hypothesis 11, which in effect is the null-hypothesis of all four previous hypotheses, assuming that *"The voting pattern of judges is not affected by the gender balance in the panel"*.

This result strongly disfavours the *Informational approach* to explaining gendered behaviour in the courtroom. In this approach, we would have to assume that men and women vote in different fashions because of their differing informational background. A logical consequence of this is that information could spread from men to women or the opposite way in deliberative forums, which a judicial panel arguably is. Since the judges' voting pattern was not affected by any measure of their peers' gender composition, we ought to assume that this has not happened. This strengthens both the *Representational* and the *Different voice* accounts of gendered behaviour.

[69] Because of the relative importance of these variables to the research question, the analysis has also been run with only one and one of these measures included. None of them proved to be significant.

If the public is a party, judges are less likely to vote in favour of the female party

In block 2, just as in block 1, none of the panel-level aggregates of significant variables from the individual level proves significant. The variables are therefore rejected, a conclusion is supported by both the LR-test and the information criteria.

Block 3 introduces three measures of case particularities, of which two are related to the gender of lawyers. All three prove not to be significant, and we reject hypothesis 14, assuming that *"Any individual gender effect is strengthened if the party which female judges are more likely to vote in favour of is represented by a female lawyer"*.

In the same block, however, we find that there is a clearly significant and seemingly strong difference in judges' voting pattern between cases where the female party is opposing a representative of the public (i.e. the childcare services) and cases where the female party opposes the child's father. As presented in this block, the odds of voting for the female party are as much as 99 % lower if the opposing party is a representative of the public interest, rather than the child's father. This is consistent with the literature on *Høgsterett*. The public interest has a generally lower overall rate of loss than private interest.

Controlling for this variable does not seem to have any particular effect on the other coefficients in the analysis, except that the individual measure of judge gender is now significant within the five per cent level. This contradicts hypothesis 15, assuming that *"Any individual gender effect will be strengthened when controlling for whether the female party's opposing party is a male individual or a representative of the public."* This conclusion is further supported by the LR test and the two information criteria.

Not surprising[70], the inclusion of this variable does have a small impact on the estimated intra-class correlation, which is reduced by a little more than one percentage points from the model of individual effects, presented in Table 5.6 on page 64. This shows that this variable controls for a small part of the level 2 variance. When the change in ICC turned out to be somewhat low, this is partly related to the low number of votes cast in cases where the opposing party is a representative of the public (35%, see section 5.1.1).

[70] See section 3.3.2 for a discussion of the ICC in logistic analyses.

The fourth variable introduced in block 3 is a control for whether there is dissent in the panel in question or not. This coefficient proves insignificant in explaining the outcome.

Time has no effect

Block 4 introduces two measures of time, none of which seems to provide a significant prediction of the outcome. This contradicts hypothesis 16, assuming that *"Any individual gender effect will be alleviated by the number of years women have served alongside men in the court"*.

This result also disfavours the informational approach. Because the informational background is what makes male and female judges act differently, we would have to assume that the more time has passed since the first woman was introduced into the judicial collegium, and the more women are present in the collegium, the more will judges of both genders have deliberated together in different panels. This ought to alleviate the differences in informational backgrounds, contrary to the assumptions made in the Representative and – especially – the Different voice accounts of gendered behaviour. Thus, the rejection of the alternative hypothesis and confirmation of the null-hypothesis is, in this case, effectively to the favour of the Representative and Different voice approaches.

Summing up the collegial effect of gender

Table 5.8: Effect of gender controlling for collegial/panel effects, but not critical actors

	coefficient	standard error	p-value	Summary
Intercept	.86	.72	.231	Level 1 units (*N*) 750 (judges)
Gender	-.93**	.43	.030	
Born in Oslo	.89***	.34	.009	Level 2 units (N) 150 (panels)
Opposing party is public	-5.62***	1.51	.000	
level 2 residual variance	34.76			LR test vs. flat logistic regression: p<.01
intra class correlation	.914			
LR-test vs. previous best model	.000			
AIC / BIC	561.8051 / 584.9055			

*Note: The dependent variable is individual judges' propensity to vote for the feminine party. There are N=750 observed votes within n=150 panels with five votes in each. The analyses are run using the xtmelogit command with 24 integration points in Stata. *** p<.01, ** p<.05, * p<.10. "(c)" indicates variable centred on mean.*

As the *Organisational* approach was clearly weakened when modelling judicial behaviour on individual level variables, the *Informational* approach has clearly been weakened when testing for collegial effects.

We find that the gender effect at the individual level is neither affected by the gender composition in the panels, nor is it alleviated over time, indicating that the presence of women in court has not led to any significant general change in practice, except in terms of the factually different behaviour of male and female judges.

5.2.3 Step 3: Identifying critical actors

The next step of the analytical strategy (see section 3.3.4) is to identify potential critical actors among the judges in the selection. To recapitulate, the purpose of this is twofold. *First*, as discussed in the methods chapter (see section 3.2.4), the chosen multi-level research design is based on a compromise due to computational limitations. There are very good reasons to assume both intra-panel dependencies *and* intra-individual dependencies. The present two-level research design, however, only accounts for the first of these two. The identification of critical actors and introduction of these into the analysis as control variables should thus introduce a certain degree of control for the last of the two types of dependencies. *Second,* although of limited immediate interest, the information gained from this part of the analysis should be of empirical interest in itself, as it sheds light on the inner workings of the judicial collegium in the court.

Potential critical actors – descriptive statistics

Table 5.9 below lists the judges in *Høgsterett* ordered by their frequency of appearances in child custody cases (between 1968 and 2011), excluding 56 judges who have appeared in fewer than ten such panels. Female judges are marked by grey background. We can see that the three female judges who have appeared in most such panels, justices Holmøy, Gjølstad and Selmer, have all appeared in more than twenty, ranging from 23 to 30 observations. Similarly, for male judges, justices Christiansen, Aasland and Sinding-Larsen are ranging from 29 to 35 observations, not surprisingly summing up to a higher share of the cases.

The vote column refers to the dependent variable in the analysis. Being a dummy, this can be interpreted as a proportion, a percentage measure, of how many times the judge in particular has voted in favour of the female party. For the 56 judges who have voted in fewer than ten

panels, the share is approximately 50% in favour of the feminine party. This diverges by only five percentage points from the general mean of 45% (see Table 5.1 on page 52).

Table 5.9: Judges in *Høgsterett* ordered by frequency of appearances in child custody cases

Name	Observations			Vote		Background		Høgsterett
	Freq.	%	Cum. %	Mean	s.d.	Birth	Birth place	Active
Arne Christiansen	35	4.67 %	4.67 %	.46	.51	1926	Oslo	1974-1996
Gunnar Aasland	33	4.40 %	9.07 %	.39	.50	1936	Bærum	1979-2006
Vera Louise Holmøy	30	4.00 %	13.07 %	.67	.48	1931	Oslo	1976-2001
Tore Sinding-Larsen	29	3.87 %	16.80 %	.62	.49	1929	Bergen	1977-1997
Liv Gjølstad	28	3.73 %	20.67 %	.29	.46	1945	Tønsberg	1988- *pres.*
Rolv Hellesylt	25	3.33 %	24.00 %	.36	.49	1927	Synnylven	1979-1997
Elisabeth S. Selmer	23	3.07 %	27.07 %	.35	.49	1923	Christiania*	1971-1990
Jan Rasmus Skåre	21	2.80 %	29.87 %	.38	.50	1929	Førde	1978-1998
Jens Bugge	21	2.80 %	32.67 %	.52	.51	1930	Oslo	1982-2000
Knut Blom	21	2.80 %	35.47 %	.57	.51	1916	Christiania*	1968-1986
Steinar Tjomsland	20	2.67 %	38.13 %	.15	.37	1948	Kristiansand	1991- *pres.*
Finn Backer	19	2.53 %	40.67 %	.47	.51	1927	Oslo	1986-1997
Karenanne Gussgard	19	2.53 %	43.20 %	.21	.42	1940	Sandefjord	1990-2010
Helge Røstad	18	2.40 %	45.60 %	.39	.50	1923	Kristiansand	1975-1993
Tore Schei†	18	2.40 %	48.00 %	.33	.49	1946	Oslo	1986-*pres.*
Lilly Helena Bølviken	17	2.27 %	50.27 %	.47	.51	1914	Arendal	1968-1984
Kirsti Coward	16	2.13 %	52.40 %	.31	.48	1940	Kristiansand	1994-2010
Hans Methlie Michelsen	14	1.87 %	54.27 %	.57	.51	1920	Bergen	1972-1990
Trond Dolva	14	1.87 %	56.13 %	.29	.47	1934	Kongsberg	1984-2004
Andreas Endresen	13	1.73 %	57.87 %	.46	.52	1908	Stavanger	1959-1978
Egil Endresen	13	1.73 %	59.60 %	.23	.44	1920	Stavanger	1977-1988
Jan Frøystein Halvorsen	13	1.73 %	61.33 %	.38	.51	1928	Oslo	1983-1995
Ketil Lund	13	1.73 %	63.07 %	.77	.44	1939	Oslo	1990-2009
Nils Peder Langvand	13	1.73 %	64.80 %	.39	.51	1929	Volda	1984-1996
Einar Løchen	11	1.47 %	66.27 %	.55	.52	1918	Bærum	1977-1985
Hans Flock	11	1.47 %	67.73 %	.46	.52	1940	Melhus	1996-2010
Magnus Aarbakke	11	1.47 %	69.20 %	.36	.51	1934	Tysnes	1994-2002
Rolv Einar Ryssdal†	11	1.47 %	70.67 %	.55	.52	1914	Laksevåg	1964-1984
Carl Ludovico Stabel	10	1.33 %	72.00 %	.33	.52	1912	Christiania*	1964-1982
Charles Philipson	10	1.33 %	73.33 %	.50	.53	1928	Oslo	1984-1990
Jens Christian Mellbye	10	1.33 %	74.67 %	.40	.52	1914	Christiania*	1968-1982
Sigurd Juell Lorentzen	10	1.33 %	76.00 %	.80	.42	1916	Sunndal	1972-1979
Trygve Leivestad	10	1.33 %	77.33 %	.30	.48	1907	Tromsø	1958-1977
56 remaining judges with <10 appearances	170	22.67 %	100 %	.51	.50			
Total	750	100 %		.45	.50			

*Note: Female judges marked by grey background colour. Only judges who have partaken in the present selection of child custody cases are included. *Christiania was the name of the capital, Oslo, until 1925. †Chief justice (justice Ryssdal served as chief justice between 1969 and 1984, justice Schei has been serving since 2002. Both former chief justices Sandene and Smith have fewer than ten observations). "Observations" refers to how many votes are observed from the relevant judge in this sample of judicial panels. "Vote" refers to the dependent variable of the analysis. Because this is a dummy variable, the mean vote can be interpreted as a percentage measure of how often the relevant judge has voted in favour of the feminine party. The total vote is the mean vote for the entire sample.*

Compared to this, one judge, justice Tjomsland, stands out with less than 20% of the votes in favour of the female party. In addition, justices Leivestad, E. Endresen, Dolva, Gussgard and

Gjølstad have all voted in favour of the feminine party in somewhere between 20% and 30% of the cases. In the other end of the scale, two judges stand out in particular. Justice Lorentzen has voted in favour of the feminine party in 80% of the cases, closely followed by justice Lund at 77%. In addition, justices Holmøy and Sinding-Larsen have both voted for the feminine party in more than 60% of the cases.

Identifying consistent voting patterns – the bivariate analyses

Table 5.10 below provides a summary of the significant results in the process of identifying critical actors in the present selection of cases[71]. As stated in the *analytical strategy* (section 3.3), any judge whose identifying dummy variable is found to predict the outcome in the dependent variable on a five per cent level, is understood to have a *consistent voting pattern*. Coefficients significant on a ten per cent level are considered marginally significant and introduced in further analysis. Notably, these analyses are bivariate, and do not control for any potential effect of other judges.

Table 5.10: Judges identified as potential critical actors

Judge	*level*	**Bivariate – consistent voting pattern**				**Trivariate – peer effect**			
		coeff.	s.e.	p-value	LR-test	coeff.	s.e.	p-value	LR-test
Leivestad	*individual*	-3.69**	1.84	0.046	**	-3.90**	1.98	0.049	
n:10	*in panel*				0.016	1.17	2.39	0.623	0.623
Tjomsland	*individual*	-2.83**	1.29	0.029	***	-2.59**	1.25	0.039	*
n: 20	*in panel*				0.010	-2.88*	1.75	0.099	0.090
Schw. Selmer	*individual*	-2.07**	1.03	0.045	**	-2.12**	1.05	0.044	
n: 23	*in panel*				.030	0.54	1.62	0.739	.739
Sinding-Larsen	*individual*	1.76*	.94	0.061	**	1.64*	.93	0.078	
n: 29	*in panel*				.046	1.47	1.49	0.322	.318
Holmøy	*individual*	1.69**	.86	.050	**	1.51*	.85	.074	*
n: 30	*in panel*				.038	2.41	1.47	.100	.091

*Note: The dependent variable is individual judges' propensity to vote for the feminine party. In the bivariate analyses, the LR-test compares with an empty, unconstrained model, while in the trivariate analyses the LR-test compares to the respective bivariate model. The analysis is run using the xtmelogit command with 24 integration points in Stata. *** p<.01, ** p<.05, * p<.10. n=the frequency of panel attendance by the judge in question.*

We can see from the results in the bivariate models that – when controlling for nothing but the chosen multi-level structure of the data – justices Leivestad, Tjomsland (male) and

[71] Only those judges who are identified as critical actors are included. A full overview, including all judges (with 10 or more observations), is presented in the appendix for reference.

Schweigaard Selmer (female) seem to have been voting significantly consistent in disfavour of the female party. The bivariate predictions range from 87% (Schw. Selmer), via 94% (Tjomsland) to 97% (Leivestad) lower odds of these judges voting for the female party than the other judges in the selection.

On the opposite, we find justice Holmøy (female), who seemingly – and controlling only for the data structure – have been voting significantly consistent in favour of the female party. In addition, we find that the pattern of justice Sinding-Larsen's (male) votes is similar to that of Holmøy, but only marginally significant, on a ten per cent level. These two judges are predicted to have between four and five times higher odds of voting for the female party than other judges in the selection.

Finally, a test analysis run with all five potential critical actors included as explanatory variables can be found in the appendix (see section 7.3.2). When modelling the dependent variable on all critical actors at once, the p-value of almost all coefficients increased, but all remain marginally significant on the ten per cent level and will be included in the further analysis.

Inter-individual dependencies revisited

Controlling only for the chosen multi-level structure of the data, the results of the bivariate critical actors analyses lend support to hypothesis 4, assuming that *"Some individual judges – so-called critical actors – will vote consistently either in favour of disfavour of the female party throughout the cases"*. This is another[72] indication that there are significant inter-individual dependencies in the present data, and that it is unreasonable to run an analysis of these data without any mechanism to control for these effects.

Critical actors diverge from the general trend

The results seem not to support to hypothesis 5 either. This hypothesis sets out that *"A large majority of the judges who are identified as critical actors will vote consistently in favour of the female party throughout the cases"*. Only one of the two female critical actors, justice Holmøy, seems to be leaning towards the female party, while the second female judge, justice Schweigaard Selmer, however, seems to lean in the opposite direction. While justice Holmøy

[72] See also sections 3.2.4 and 3.3.2.

is accompanied by male critical actor justice Sinding-Larsen, justice Schweigaard Selmer is accompanied by both justices Leivestad and Tjomsland.

This divergence within the gender groups for the critical actors also contradicts hypothesis 6, assuming that *"Any judges identified as critical actors will vote consistently in favour of their gender group's preferred party"*. Of the female judges, only Schweigaard Selmer votes in agreement with the gender effect found in the previous models of gendered behaviour, where female judges are less likely to vote for the female party than male judges are. Justice Holmøy, on the other hand, stands out as an anti-woman, with a voting pattern contrary to the predicted pattern of her gender group. With regard to the three male critical actors, we see the same pattern, where two of three, justices Leivestad's and Tjomsland's predicted voting pattern has the opposite direction of their gender group's voting pattern. These results disfavour the *Different voice* approach, in which we should assume that male and female perspectives are fundamentally different.

Notably, any final conclusions with regard to these hypotheses are dependent on the outcome of step four of the analysis, *adding the effect of critical actors to the gender model*.

Some statistical reservations with regard to the critical actors analysis

The similar direction of effect in Leivestad's, Tjomsland's and Schweigaard Selmer's coefficients could be due to a high degree of "panel overlap". These three judges could have coincidentally partaken in several of the same panels, in which the case particularities were so clear that the outcome was given. However, Schweigaard Selmer and Leivestad only served together in two panels (Rt-1971-1253 and Rt-1976-644), while the combinations Schweigaard Selmer and Tjomsland, and Tjomsland and Leivestad never occur in the same panel. Seeing that all of these judges have served in at least ten panels (excepting Tjomsland, at least twenty, see Table 5.9 above), this is not a high overlap. Similarly, we can count three cases in which both justices Holmøy and Sinding-Larsen partake in the same panel (Rt-1984-728, Rt-1985-179 and Rt-1987-598). This does not constitute a high overlap either.

Additionally, because dissenting votes stand out as variation within panels, we could imagine that the bivariate analyses in this test mainly picks out only those judges with high dissent rates, i.e. votes in opposition to their peers in panel. This would *exclude* any judges who actually affect their peers' voting pattern and do not need to dissent. As presented in section

5.2.1, the average rate of dissent against the majority vote in this selection is 8%. Compared to this, three of the identified critical actors do have somewhat higher dissent rates: As much as 20% of justice Leivestad's votes; 17% of justice Schweigaard Selmer's votes; 13% of justice Holmøy's votes are dissenting votes. However, justice Tjomsland's dissenting votes count 10%, only slightly above the mean, while justice Sinding-Larsen's dissenting votes count 7%, slightly below. This gives an average dissent rate for the five critical actors of 11%, only 3 percentage points higher than the general mean.

Identifying peer effects – the trivariate analyses

The rightmost column of Table 5.10 shows us that, of the five justices identified as critical actors in the bivariate analyses, only in the case of Tjomsland does the aggregate "Tjomsland in panel" variable show a marginally significant effect on the outcome. In the case of Holmøy, however, as with Tjomsland, the LR-test also shows marginally significant improvement of the model's explanatory power when including the aggregate variable. With a preferred significance level of five per cent, these results lend little support to hypothesis 13, assuming that *"The presence of a critical actor in a panel significantly raises the likelihood of a judge voting in favour of the critical actor's preferred party"*. This disfavours the Informational approach, in which we should assume that a critical actor would be able to provide unique information to any deliberative panel about his or her attitudes to a matter, thus also affecting her or his peers' attitudes.

The general impression from the final part of identifying and measuring the effect of critical actors on their peers is that however much some judges seem to vote significantly consistent, this is mainly an expression of individual particularity. The critical actors identified here do not seem to affect their peers' voting pattern in any significant manner, but rather show a slight tendency of divergence.

The next step of the analytical strategy is to add the effects of the identified critical actors to the gender model as a control for *inter-individual dependencies*.

5.2.4 Step 4: Adding the effect of critical actors to the gender model

Table 5.11 Exploring the impact of critical actors

Variables	**Block 1** **Critical actors individually**			**Block 2** **Critical actors peer effects**			**Block 3** **Summing up: the significant effects**		
	coeff.	s.e.	p	coeff.	s.e.	p	coeff.	s.e.	p
Intercept	1.12	.81	.167	1.09	.84	.193	1.40*	.72	.053
Gender	-1.32**	.61	.031	-1.76***	.51	.001	-1.77***	.51	.001
Born in Oslo	.72*	.39	.069						
Opposing party is public	-6.24***	1.67	.000	-5.41***	1.52	<.001	-5.55***	1.38	<.001
Justice Leivestad (m)	-4.19**	2.12	.048	-4.16**	2.00	.038	-4.30**	2.00	.031
Justice Tjomsland (m)	-2.38*	1.30	.067	-2.74**	1.26	.030	-2.74**	1.26	.030
Justice Schweigaard Selmer (f)	-1.24	1.29	.337						
Justice Sinding-Larsen (m)	1.83*	1.04	.080						
Justice Holmøy (f)	2.31**	1.10	.036	2.92***	.99	.003	3.04***	.99	.002
Justice Leivestad in panel				-.64	2.37	.787			
Justice Tjomsland in panel				-.13	1.94	.948			
Justice Holmøy in panel				1.57	1.46	.284			
Random effects									
Level 2 residual variance	42.56			35.66			36.01		
Intra-class correlation	.928			.916			.916		
LR-test	<.001[73]			.726[74]			<.001[75]		
AIC *BIC*	547.9386	594.1394		553.3541	599.5548		548.6695	581.0101	

*Note: The dependent variable is individual judges' propensity to vote for the feminine party. There are N=750 observed votes within n=150 panels with five votes in each. The analyses are run using the xtmelogit command with 24 integration points in Stata. *** p<.01, ** p<.05, * p<.10. "(c)" indicates variable centred at mean. LR-tests compare with the previous block which significantly contributed to the explanatory power of the model. In addition, where the entire block itself does not contribute to the explanatory power of the model, but one or some of the explanatory variables do and are included in the following blocks, the LR-test compares with downscaled versions of the block, because comparison to a rejected model is meaningless. Stata output for the downscaled versions can be found in the appendix for reference.*

[73] LR-test against the best model of both individual and collegial controls on the gender effect (see page 55)

[74] LR-test against the best possible combination of variables from block 1 in this table (excluding *Born in Oslo*, *Justice Selmer* and *Justice Sinding-Larsen* from the equation).

[75] LR-test against the best model of both individual and collegial controls on the gender effect (see page 55). An LR-test comparing block 1 and three in this table automatically assumes that the third block is nested in the first because of the number of variables. This shows that the increase in explanatory power for block 1 compared to block three is only marginally significant (p=.081), favouring a rejection of block 1 for block 3.

Exploring the impact of potential critical actors

The first block of Table 5.11 is a presentation of the gender model when controlling for the individual-level dummy variables identifying critical actors. The second block introduces the aggregate variables for those critical actors that are found to be significant on the basis of block 1 (this is further elaborated upon below). Block 3 is a presentation of the end product of this analysis, summing up all those effects found to be significant in the previous two blocks.

All explanatory variables from the temporarily final model in step two of the analysis (see page 69) are reintroduced in this step of analysis, and the LR-test in both blocks 1 and 3 compares to this temporarily final model. The main interest of this step is to examine to which degree inter-individual dependencies function as a bias in the analyses. First, each coefficient will be interpreted and investigated, before the effect on other coefficients can be analysed. The latter happens mainly on basis of block 3.

Interpreting the critical actors' coefficients

Block 1 in the table confirms justices Leivestad's and Holmøy's significantly consistent voting pattern. These are included in the further analysis. Further, when controlling for gender, geography and the role of the opposing party, Schweigaard Selmer's voting pattern no longer seems significantly consistent, and is excluded from further analysis.

Although the direction of Justices Tjomsland's and Sinding-Larsen's coefficients is still the same, when controlling for the significant effects from previous parts of the analysis, these are now only marginally significant and fosters further investigation. Stata outputs of the stepwise removal of explanatory variables from block 1 to block 2 in the table above can be found in the appendix (see section 7.8). The model has been gradually reduced by first removing Schweigaard Selmer, continuing with the Oslo-variable[76], Sinding-Larsen and, in the end Tjomsland. Each reduced block was tested by LR-tests, concluding that controlling for neither Schweigaard Selmer, the Oslo effect, nor Sinding-Larsen contributes significantly to the explanatory power of the model as a whole. In the opposite direction, the model including a control for the Tjomsland variable seems to have significantly improved explanatory power compared to a reduced model without the Tjomsland variable, indicating that this coefficient

[76] The Oslo effect is further discussed in the following section.

should be refined in further analysis. This leaves us with justices Leivestad, Holmøy and Tjomsland as controls for inter-individual dependencies.

Unsurprisingly, introducing the aggregate critical actors' variables in block 2 provides little improvement. In accordance with the previous step of the analysis, none of the coefficients are significant or contribute to the model, a conclusion supported by the LR-test.

Block 3 is the summary of the two previous blocks, including only the explanatory variables found to have a significant effect on the outcome. This forms the basis for the following step of analysis, which is to examine whether inter-individual dependencies function as bias in the analyses.

The Oslo effect is a spurious critical actors effect

First, the coefficient for the geography variable is fairly similar from Table 5.8 to block 1 in Table 5.11 above. When controlling for critical actors, however, geography has gone from being clearly significant on a one per cent level to being only marginally significant, raising its p-value almost seventyfold. Seemingly, this is a spurious effect, where diverse inter-individual dependencies coincidentally have been lying in the background dragging this coefficient up.

To explain this, we need to have a look at the critical actors themselves: On one side, justice Holmøy, who tends to vote in disfavour of the feminine party, is Oslo-born and has boosted the "Oslo effect". On the other side, justices Tjomsland and Leivestad are both born outside of Oslo and tend to vote in disfavour of the feminine party. Thus, they have similarly contributed to the negative effect of not being from Oslo. All together, these three justices have voted 60 times in this selection of cases, enough to have a strong influence on the Oslo variable. Thus, methodologically, the observation that only three strong individuals – *critical actors* – can account for an entire spurious effect illustrates the importance of controlling for inter-individual dependencies in future analyses of judicial behaviour. In this case, a potential type II error has been avoided.

The gender effect is strengthened

With regard to the gender coefficient, we see that this is both severely strengthened and has gone from being marginally significant (p=.053) before controlling for critical actors, to being clearly significant on a one per cent level in block 3 of the table above. This contradicts

hypothesis 12, assuming that *"Parts of any individual gendered voting pattern found in the analysis can be explained by the presence of a critical actor in the panel"*. To reject the null-hypothesis in this regard we would have to see a significant *weakening* of the gender effect, not a strengthening. Similar to several results before, this disfavours the informational approach, which assumes that information contributed from critical actors in deliberative panels would affect other judges' voting pattern.

The fact that gender comes out as a strengthened predictor of individual judges' voting pattern can be explained in a similar fashion as with the spurious Oslo effect. Justice Holmøy, a female judge, has been voting consistently *in favour of* the feminine party, while justices Tjomsland and Leivestad, both male judges, has been voting consistently *in disfavour* of the feminine party. In a manner of speaking, we can say that Holmøy has been voting as an *anti-woman*, while Tjomsland and Leivestad have been voting as *anti-men*. All three have thus significantly weakened the general effect of gender, another conclusion illustrating the importance of testing for inter-individual dependencies. In this case, a potential type I error has been avoided.

5.2.5 Steps 5 and 6: Varying slopes and cross-level interactions

So far, only random intercept models have been run. The next step of analysis is a random coefficient model, as recommended by Hox (2010) and according to the analytical strategy presented in sections 3.3.6 and 3.3.7. Because this is a very demanding process (Hox, 2010:58), one and one coefficient is tested for varying slopes and compared to the previous block with an LR-test. Stata outputs of this process can be found in the appendix (see section 7.9). They are not included here because none of the slopes proved to have any significant variance components between the panels. Because this process forms the basis for the last step of the analytical strategy, there is no basis to move forward to the last step of the analysis, examining cross-level interactions.

5.2.6 The final model of gendered behaviour in *Høgsterett*

Table 5.12: The final model

	coefficient	standard error	p-value	Summary
Intercept	1.40*	.72	.053	
Gender	-1.77***	.51	.001	Level 1 units (*N*) 750 (judges)
Opposing party is public	-5.55***	1.38	<.001	
Justice Leivestad	-4.30**	2.00	.031	Level 2 units (N) 150 (panels)
Justice Tjomsland	-2.74**	1.26	.030	
Justice Holmøy	3.04***	.99	.002	
level 2 residual variance		36.01		
intra class correlation		.916		LR test vs. flat logistic regression: p<.01
LR-test vs. previous best model		<.001		
AIC / BIC		548.6695 / 581.0101		

*Note: The dependent variable is individual judges' propensity to vote for the feminine party. There are N=750 observed votes within n=150 panels with five votes in each. The analyses are run using the xtmelogit command with 24 integration points in Stata. *** p<.01, ** p<.05, * p<.10. "(c)" indicates variable centred on mean.*

Table 5.12 is a presentation of the end-result of the analysis and, because step 5 and 6 of the analytical strategy proved unfruitful, this is equal to block 3 of Table 5.11, which summed up the effects of adding controls for critical actors to the analysis.

Studying the final model, we see that female judges in *Høgsterett* are significantly less likely to vote for the female party in child custody cases than are male judges. Controlling for the structure of the data, the nature of the opposing party and three critical actors, we find that the odds of female judge voting in favour of the female party is as much as 83% lower than that of a male judge[77]. Additionally, and with the same constraints, we see that the odds of a judge – regardless of gender – voting in favour of the female party is as much as 99% lower if the opposing party is a representative of the public, rather than a male individual.

Male and female judges are likely to exercise discretion differently

When interpreting these results, however, we should keep in mind that this analysis controls for the panelled structure of the data. This means that calculated odds cannot be interpreted as straightforward as in a flat logistic analysis. The intra-class correlation (ICC) has remained

[77] See footnote 64 on page 61 for an explanation of how this is calculated.

quite stable at approximately 92% during the entire analysis[78]. As discussed at the first step of the analysis (section 5.2.1), this leaves only 8% of the variance in the outcome to be explained by variables on the individual level. Variables measuring collegial factors related to gender, geographical background and experience in Table 5.7 did not turn out to have any significant effect on the outcome. The only variable on the collegial level to show significant effect on the outcome is a measure of case particularities, the nature of the opposing party, but adding this variable did not have a large impact on the ICC.

What this shows us is that this model explains just a fraction of the variance in the dependent variable, and we should therefore not assume that this is a good model to predict the outcome of child custody cases, or even to predict the voting pattern of judges within this category of cases.

What it does tell us something about is the nature of the impact of gender on judicial behaviour: In this respect, it gives us a clear indication that when judges exercise discretion, female and male judges are likely to end up with quite different conclusions.

To gain insight into why this is the case, we need to evaluate to which degree any of the four general approaches to gendered behaviour seem to be strengthened by the results of the analysis. The following section is an evaluation of the four approaches in lieu of the empirical analysis.

[78] We should have expected that adding the effect of having a public party rather than the child father as opponent would lower the ICC.

Table 5.13: Summary of the hypothesis evaluation

Theory	Expected effect	Actual effect	Evaluation
Expected individual effects			
Different voice, Informational, Representational *Disfavouring:* Organisational	*H1: "Male and female judges vote significantly different from each other"*	Gender is a significant factor in predicting judges' votes.	Hypothesis confirmed
Representational *Disfavouring:* Organisational	*H2: "Female judges are significantly more likely than male judges to vote in favour of the female party"*	The effect of gender goes in the opposite direction.	Hypothesis rejected.
Organisational *Disfavouring:* Different voice, Representational, Informational.	*H3: "Any difference found in the voting pattern of male and female judges is purely random"*	Gender is a significant factor in predicting judges' votes.	Hypothesis rejected.
Informational, Representational *Disfavouring:* Organisational	*H4: "Some individual judges – so-called critical actors – will vote consistently either in favour or disfavour of the female party throughout all cases"*	Several critical actors were identified.	Hypothesis confirmed
Representational *Disfavouring:* Organisational	*H5: "A large majority of the judges who are identified as critical actors will vote consistently in favour of the female party throughout the cases"*	Critical actors voted in both ways within each gender group. In the final model, only one of three voted for the female party. This was also the only woman.	Hypothesis rejected
Different voice *Disfavouring:*	*H6: "Any judges identified as critical actors will vote consistently in favour of their gender group's preferred party"*	Critical actors voted in both ways regardless of gender	Hypothesis rejected.
Expected collegial/panel effects			
Informational, Representational, *Disfavouring:* Different Voice	*H7: "Any individual gender effect will be strengthened in panels with a critical mass – a majority – of women."*	No measures of gender on the panel level were significant.	Hypothesis rejected.
Informational *Disfavouring:* Representational, Different voice, Organisational	*H8: "Controlling for any individual effect of gender, the voting pattern of judges is affected by the gender balance of the panel in which they vote."*	No measures of gender on the panel level were significant.	Hypothesis rejected.

Informational *Disfavouring:* Representational, Different voice, Organisational	*H9: "Having at least one woman present will affect the voting behaviour of all judges in the panel."*	No measures of gender on the panel level were significant.	Hypothesis rejected.
Informational *Disfavouring:* Representational, Different voice, Organisational	*H10: "Controlling for any individual effect of gender, the gender of the presiding judge will affect the voting pattern of other judges in the panel."*	No measures of gender on the panel level were significant.	Hypothesis rejected.
Organisational, Representational Different voice *Disfavouring:* Informational.	*H11: "The voting pattern of judges is not affected by the gender balance in the panel"*	No measures of gender on the panel level were significant.	Hypothesis confirmed.
Informational *Disfavouring:* Representational, Organisational, Different voice	*H12: "Parts of any individual gendered voting pattern found in the analysis can be explained by the presence of a critical actor in the panel in which they cast their vote."*	Controlling for critical actors strengthened the gender effect. We should have expected the opposite effect.	Hypothesis rejected.
Informational *Disfavouring:* Different voice, representational, organisational	*H13: "The presence of a critical actor in a panel significantly raises the likelihood of a judge voting in favour of the critical actor's preferred party."*	None of the critical actors significantly affected their peers' voting pattern.	Hypothesis rejected.
Expected effects of case particularities			
Different voice	*H14: "Any individual gender effect is strengthened if the party which female judges are more likely to vote in favour of is represented by a female lawyer"*	The gender of lawyers does not have significant effect on judges' votes.	Hypothesis rejected.
Representational	*H15: "Any individual gender effect will be strengthened when controlling for whether the female party's opposing party is a male individual or a representative of the public."*	Controlling for who the opposing party represents had no significant effect on the gender coefficient.	Hypothesis rejected.
Informational *Disfavouring:* Representational, Different voice	*H16: "Any individual gender effect will be alleviated by the number of years women have served alongside men in the court"*	None of the time measures were significant.	Hypothesis rejected.

Note: This is summary of the hypothesis evaluation conducted section 5.2, as part of the analysis.

5.3 Evaluating the four approaches

The table below summarises the evaluation of hypotheses conducted throughout the analysis section. In the following section, each theoretical approach's ability to explain the observed gendered behaviour is evaluated based on the outcome of the hypothesis tests.

Both the Organisational and the Informational approaches are rejected based on the hypothesis tests, and these two are evaluated first. After this, the weaknesses and strengths of the two most promising approaches, the Representational and the Different Voice approaches, are evaluated. The implications of the present evaluation are presented in the next and final chapter of the thesis.

5.3.1 The Organisational approach is weakened

Table 5.13 provides a summary of the hypothesis evaluation conducted throughout the analysis chapter. The first hypothesis to be tested was the suggestion that *"male and female judges vote significantly different from each other" (H1).* Because the Organisational approach assumes that gender has no influence on the behaviour of judges, a confirmation of this hypothesis clearly disfavours this approach. It would also be an automatic dismissal of H3: *"Any difference found in the voting pattern of male and female judges is purely random"*, which is inferred from the Organisational approach and in effect is the null-hypothesis of H1.

The confirmation of H1 and dismissal of H3 clearly weakens the Organisational approach's rejection of gendered behaviour. The present analysis shows that male and female judges do have significantly different voting patterns. The confirmation of hypothesis 4, that *"Some individual judges – so-called critical actors – will vote consistently either in favour or disfavour of the female party throughout all cases"*, further weakens the Organisational approach. In combination, these two findings give a clear impression that judges are more than their education and professional background.

However, although the Organisational approach seems to fail in its assumptions on gendered behaviour, we should not reject it all together. The final gender model in the analysis seems to explain only a fraction of the outcome in the dependent variable. By far, most of the variation in the judicial votes is explained by either collegial factors or case particularities which the gender approaches in this analysis do not (and does not intend to) touch. This accords with the most recent statistical analyses of judicial behaviour in *Høgsterett*, concluding with the need

for a stronger focus on case particularities in future analyses (see Bentsen (2012:88), Skiple (2012:87) and Jacobsen (2012:93)). Similarly, Sunde (2012a:196-7) has called for the addition of measures of how judges make use of the sources of law. It also accords with the international development of the field of judicial behaviour, as scholars are increasingly calling for models that incorporate a wider range of explanatory factors (see e.g. Epstein and Knight (2013:25)).

5.3.2 The Informational approach does not explain the gender effects

The Informational approach is the only approach to gendered behaviour within which we can expect general collegial effects of gender. The hypotheses 7, 8, 9 and 10 (see Table 5.13), inferred from the Informational approach, all assume that a panel's gender balance – in some form – affects the votes of its judges. Hypothesis 7 set out that serving on a panel with a critical mass of female judges would change the gender effect for judges. Hypothesis 8 is the assumption that the gender balance in itself affects the vote. Hypothesis 9 suggests that having just one woman present affects the vote of the panel peers, and, finally; hypothesis 10 is the assumption that a female presiding judge can affect the votes of her peers.

All of these hypotheses were rejected through the empirical analysis. The only significant effect found on the collegial level was a measure of case particularity – whether the opposing party is a representative of the public or a private party – and not a collegial effect per se. Consequently, hypothesis 11, which in effect is the null-hypothesis of the three aforementioned hypotheses, assuming that *"The voting pattern of judges is not affected by the gender balance in the panel"* was confirmed.

No critical actors were found to have any collegial effects on their peers' votes. What should be noted, however is that controlling for the individual effect of critical actors did strengthened the gender effect. The hypothesis inferred from the Informational approach, however, was an assumption of the opposite effect. Furthermore, hypothesis 16, assuming that *"Any individual gender effect will be alleviated by the number of years women have served alongside men in the court"*, was also rejected. None of the two measures of time have any effect on the voting pattern of judges in child custody cases.

As a conclusion, although female and male judges generally seem to exercise discretion in differing manner, this analysis gives no indication of any collegial effects of gender. The

results of the empirical analysis clearly contradict the explanations inferred from the Informational approach, within which we need to assume that gender differences are rooted in the two gender groups' differing informational background. Female and male judges are assumed to bring different general experiences with them into their profession, and this affects their reasoning and voting behaviour. Through deliberation in the judicial panel judges should thus be able to affect each other's reasoning and voting behaviour. This does not seem to be a good explanation of gendered behaviour in child custody cases before the Supreme Court of Norway.

This conclusion is added to a line of mutually contradicting analyses of gendered judicial behaviour in different contexts. For example, it contradicts Boyd et al.'s (2010) results from the US context, but concurs with the findings of Ostberg and Wetstein (2007) who found "no evidence that justices from the opposite sex moved toward each other's ideological positions" in the Canadian Supreme Court.

5.3.3 The Representative and Different Voice approaches

The final two approaches to gendered behaviour are the Different Voice approach and the Representative approach. Both approaches assume differences between male and female judges and none of the two assume any collegial gender effects. As such, they fare better in explaining the gendered behaviour in this category of cases than the two previously evaluated approaches. On the other hand, both of these seem to have weaknesses in explaining the gendered behaviour of judges in child custody cases.

Evaluating the Different Voice approach

Where the Representational approach assumes that female judges (and perhaps a few male judges) act as representatives of a perceived "female class", the Different Voice approach assumes that male and female reasoning is fundamentally different altogether. Hypothesis 6 inferred from the Different Voice approach suggest that "*Any judges identified as critical actors will vote consistently in favour of their gender group's preferred party*". This is reasonable because female judges should be reasoning in a female manner if the Different Voice assumption was true. What the analysis showed, however, is that the two female judges identified as critical actors in section 5.2.3, voted in both ways regardless of gender. Additionally, all three critical actors who were included in the final model voted in dis-accordance with their gender group association. Clearly, individual exceptions to any norm exist, and we should also expect that in this case. These are but a few individuals, but the

results do indicate that gender differences are not as deeply rooted as would be assumed with a Different Voice approach.

An additional assumption made particularly on behalf of the Different Voice approach is that *"Any individual gender effect is strengthened if the party which female judges are more likely to vote in favour of is represented by a female lawyer"* (*H14)*. This is based on a similar argument as above: If male and female reasoning differs in a fundamental manner, we can imagine that women are better equipped to convince other women than other men. Additionally, because female lawyers before *Høgsterett* are still the anomaly, we can assume that any differences between judges are strengthened particularly when a female lawyer is present.

The hypothesis was rejected. The analysis gives no indication that judges are affected by the gender of the lawyers before the court.

Evaluating the Representational approach

The Representational approach is the only approach from which a particular direction of the gender effect was assumed in the theory section: *"Female judges are significantly more likely than male judges to vote in favour of the female party" (H2)*. The rationale behind this hypothesis is that a representative of the interests of women should vote in favour of the woman in child custody, rather than e.g. the farther or the public interest. Partly as a logical consequence of this assumption, hypothesis 5 was also inferred from the Representative approach: *"A large majority of the judges who are identified as critical actors will vote consistently in favour of the female party throughout the cases"*. Critical actors in the Representational approach are understood to be "more class-conscious" than other actors. Because the critical actor's by definition are diverging from the norm of the court, and because we should assume that the "male paradigm" is the norm, most critical actors' divergence is likely to favour the female interest – regardless of the actors' own gender.

Both of these two hypotheses were rejected, weakening the Representational approach. A review the theoretical basis for these hypotheses could be fruitful, however[79]. The assumption made on behalf of the Representational approach is that voting in favour of "the female

[79] This moderation was suggested during a presentation of the analysis before a professional audience of researchers.

interest" is equal to voting in favour the female party in a trial. Arguably, this assumption is narrower than need be. Behuniak-Long (1992), criticising Sherry's (1986) seminal adaption of the Different Voice theory to the study of jurisprudence, separates between the two concepts "feminist jurisprudence" and "feminine jurisprudence". In her study of Sandra Day O'Connor's behaviour as the first female judge at the US Supreme Court, she argues that O'Connor's jurisprudence could be both feminine and "deleterious to feminists". With this criticism in mind, if we assume that female judges act as representatives of the female class, it is reasonable to assume that they vote in a *feminist* way. We can further assume that feminist jurisprudence is oriented towards a gender role pattern that is different from the traditional, and that a traditional gender role pattern is one where the best interest of a child is to be with its mother. As such, we could equally well infer from the Representative approach a hypothesis saying that female judges could be prone to vote in favour of the non-feminine party.

As was mentioned under the evaluation of the Informational approach, serving on a panel with a critical mass of female judges does not seem to change the gender effect for judges in any way, leading to the rejection of hypothesis 7. The rationale for this hypothesis in the Representative approach is not one of direct collegial effects of gender, but that judges are rational actors who might abstain from voting for their preferred party unless they are in a favourable situation (see section 2.3.2).

Furthermore, hypothesis 15, assuming that *"Any individual gender effect will be strengthened when controlling for whether the female party's opposing party is a male individual or a representative of the public"* was also rejected. The nature of the opposing party has a very strong and significant effect on the voting pattern of judges, but the effect of gender is similar both before and after controlling for this factor.

Finally, the confirmation of hypothesis 4, resulting from the identification of several critical actors fits well with the Representational approach, although the direction in which the actors voted is ill-fit with the assumptions in the theory chapter.

6 Implications for our understanding of gendered behaviour

The introduction to this thesis was opened with a reference to journalist Inge D. Hansen's (2010) summary of the process of a disputed child custody case after the split between a young couple with a child in common. Hanssen had observed that on every court level, this case had a different outcome, and on every level the outcome of the case seemed to correlate with the gender of the judges who cast the verdicts. One female district judge and a panel of only female Supreme Court judges (the only in this selection) voted in favour of the child mother. In the middle level, a panel of three male appellate judges found for the father. Hanssen's open question question was: could this be a coincidence?

The purpose of this thesis was *to examine how gender affects the voting pattern of Norwegian Supreme Court judges in cases of disputed child custody*, and it provides Hanssen with an answer[80]: It is a well-known story that humans are better at recognising patterns than recognising that the patterns are in fact coincidental. This seems to be the case also with Hanssen's observation. The present analysis shows that, in child custody cases, female Supreme Court judges *do* tend to vote somewhat different from male judges in the same court. The tendency in the two gender groups is, however, opposite from the effect Hanssen meant to have seen. In addition, gender seems to explain only a fraction of why judges vote the way they do. As such, gender alone would be a very poor tool to predict how judges would vote in child custody cases. In an elaborate answer to Hanssen's question, we could also add that the difference between male and female judges is likely to be best explained either in a class-perspective or as an expression of fundamental differences in how men and women reason.

This chapter summarises the results and lists the main implications of the analysis. The implications are split into three different categories: empirical, methodological, and theoretical implications. Suggestions for further research are given as part of the implications.

[80] This is an answer with limitations: The present analysis is conducted merely on judicial votes cast in the Supreme Court, and we should be careful with assuming that conclusions in this analysis can be generalised to the lower courts. The indications from the analysis is, however, clear.

6.1 Empirical implications

The thesis' analysis has led to three main empirical observations:

1) There is a significant difference in the voting behaviour of male and female judges in child custody cases.
2) Some judges, so-called critical actors, exercise voting patterns that differs significantly from the norm of the collegium.
3) Seemingly, most of the variation in judges' voting pattern is related to particularities of the cases.

1) Gender affects the judges of the Supreme Court of Norway

Male and female justices do have voting patterns that significantly differ. Observing the final model of the analysis, we can calculate that the odds of a male *Høgsterett* judge voting in favour of the female party in a case of disputed child custody is 83% higher than for a female judge[81]. This result is strongly moderated by the fact that only 8% of the variation in the judicial vote is explainable by individual factors. We should therefore assume that the difference is more prominent in e.g. cases of dissent and when exercising judicial discretion.

Controlling for a range of factors has not identified this to be a spurious effect. Rather, when adding the effect of critical actors to the model, gender stood out as a stronger explanatory variable than before.

2) Critical actors

One of the major findings in this analysis is the identification of critical actors who seem to diverge from the norm of the collegium by voting consistently either in favour or disfavour of the female party. The identification of critical actors in this analysis is interesting in itself. The present analysis provides little information about why theses judges stand out as critical actors, however. What makes justices Holmøy, Tjomsland and Leivestad diverge from the common pattern?

A potential analysis of that question could be modelled on Skivenes' (2010) Habermasian discursive analysis of three child custody cases before *Høgsterett*. Not focusing on how the legal method is implemented, Skivenes' focus is on the logical consistency of the judges'

[81] Controlled for the multi-level structure of the data, various intra-individual dependencies and for whether the opposing party is a man or a representative of the public interest.

argumentation, finding clear variance between the selected cases, both as to which arguments were included, and as to the quality of the justification of these arguments. Implementing this method to analyse the argumentation in child custody cases where the three critical actors were first-voters (i.e. writers of the judicial argument), could provide interesting insights into the behaviour of judges.

3) Particularities of the cases seems to be important explanatory factors

This analysis explains only a fraction of what makes judges vote in the manner that they do. In addition, it provides indications that a large part of the explanations is to be found in hitherto unmeasured factors related to case particularities. In this sense, the analysis accords with recent analyses of judicial behaviour in *Høgsterett* (see Bentsen (2012:88), Skiple (2012:87) and Jacobsen (2012:93)) and legal scholars (e.g. Sunde (2012a:196-7)) who has called for measures judges' usage of the sources of law. See also Epstein and Knight (2013:25).

6.2 Methodological implications

The importance of inter-individual dependencies

Not only do some individual judges stand out as critical actors - which is interesting in itself – but adding these as control variables in the analysis is a seemingly prudent way to control for intra-individual dependencies.

In addition, the analysis has clearly shown that failing to take into account *inter-individual dependencies* when modelling judicial behaviour, means risking both type I and II errors (see section 5.2.4).

6.3 Theoretical implications

In a review of the four contending approaches to gendered judicial behaviour, the *Organisational approach*, assuming no effect, neither related to gender nor to critical actors, is considerably weakened. Similarly, the *Informational approach*, assuming collegial effects of gender is weakened. Although gendered behaviour does seem to occur on the individual level, no evidence of such can be found on the collegial level. In fact, the only significant panel level variable in the final model is the control for whether the opponent represents the public interest. That variable does not measure any collegial effects, but case particularity.

We are then left with two options, the Different Voice approach and the Representational approach, of which both correctly predicted a significant difference in voting pattern between male and female judges individually. A major difference between the two is, however, that the latter assumes critical actors, while the first only "tolerate" critical actors insofar as they vote in a manner similar to the general pattern of their gender group. All three critical actors in the thesis' final model vote in discordance with the general gender trend.

On the other side, where the Different Voice approach provides no indication as to the direction of the gender effect, the Representative approach is understood in this thesis to assume a support for one's own gender. The empirical effect is opposite: male justices are more likely to favour the female party. As such, both the Different Voice – and the Representational approaches seem to have some weaknesses. However, with a moderation of the theoretical perspective assumed on behalf of the Representational approach, the latter seems to provide a fairly good explanation of gendered voting behaviour at the court.

Expanding this analysis from studying a single category of cases to several categories could be fruitful in separating between these two approaches. The Different voice approach assumes gendered effects to be found within *all* types of cases, while the Representational approach restrict these effects to *only* gender-related issues. An analysis including – and controlling for – both gender-related and non-gender-related issues would therefore effectively separate between the two approaches.

References

Acock, A. C. (2010) *A Gentle Introduction to Stata,* College Station, Texas, Stata Press.

Allen, David W. & Wall, Diane E. (1993) Role Orientations and Women State Supreme Court Justices. *Judicature,* 77**,** 156-65.

Aschehough, Marit (2011) *List: Advokater med møterett - medlemmer av Advokatforeningen*. Available from http://www.advokatbladet.no/Nyheter1/Hoyesterettssakene-2011/. [Accessed on February 15th, 2012. Cached copy available on demand if depublished.].

Aschehough, Marit (2012) Høyesterettsdommere drikker te og blogger ikke. *Advokatbladet.* Oslo.

Baldez, Lisa, Epstein, Lee & Martin, Andrew D. (2006) Does the U.S. Constitution Need an Equal Rights Amendment? *The Journal of Legal Studies,* 35**,** 243-283.

Baum, Christopher F. (2006) *An Introduction to Modern Econometrics Using Stata,* College Station, Texas, Stata Press.

Behuniak-Long, Susan (1992) Justice Sandra Day O'Connor and the Power of Maternal Legal Thinking. *The Review of Politics,* 54**,** 417-444.

Bentsen, Henrik Litlerlé (2012) «Vår evne til å kompromisse» - Flernivåanalyse av årsakene til dissens i straffesaker i Norges Høyesterett fra 1996 til 2011. *Department of Comparative Politics.* University of Bergen.

Bergset, Kurt-Rune (2011) Report: Expanding DORANOH June-August 2011 [unpublished report]. Judicial Behaviour on the Supreme Court of Norway and the University of Bergen.

Boyd, Christina L., Epstein, Lee & Martin, Andrew D. (2010) Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science,* 54**,** 389-411.

Bull, Kirsti Strøm (2001) *Nye høyesterettsdommer om samiske rettigheter*.

Cameron, Charles & Cummings, Craig (2003) Diversity and Judicial Decision Making: Evidence from Affirmative Action Cases in the Federal Courts of Appeals, 1971-1999. *The Crafting and Operating Institutions conference.* Yale University.

Centre for Multilevel Modelling (2013) *LEMMA (Learning environment for multilevel methodology and applications)* Bristol, University of Bristol. Available from http://www.cmm.bris.ac.uk/lemma/. [Accessed on 14th of April, 2013. Cached copy available on demand if depublished.].

Davis, Sue (1994) Do Women Judges Speak 'In a Different Voice?' Carol Gilligan, Feminist Legal Theory, and the Ninth Circuit. *Wisconsin Women's Law Journal*, 8**,** 143-73.

Drammens tidende (2009) Eldre dommere gir barna til mor. *Drammens tidende.* Drammen.

Dyevre, Arthur (2010) Unifying the field of comparative judicial politics: towards a general theory of judicial behaviour. *European Political Science Review,* 2**,** 297-327.

Eckhoff, Torstein & Helgesen, Jan E. (1997) *Rettskildelære (e-book, Nasjonalbiblioteket),* Oslo, Tano Aschehoug.

Eikemo, Terje Andreas & Clausen, Tommy Høyvarde (Eds.) (2007) *Kvantitativ analyse med SPSS,* Trondheim, Tapir akademisk forlag.

Epstein, Lee & Knight, Jack (2013) Reconsidering Judicial Preferences. *Annual Review of Political Science,* 16**,** 11-31.

Farhang, Sean & Wawro, Gregory (2004) Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision Making. *Journal of Law, Economics, and Organization,* 20**,** 299-330.

Field, Andy (2009) *Discovering statistics using SPSS : (and sex and drugs and rock 'n' roll),* Los Angeles, SAGE.

Fielding, Antony & Goldstein, Harvey (2006) *Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review* Nottingham (UK), DfES Publications.

George, Alexander L. & Bennet, Andrew (2005) *Case Studies and Theory Development in the Social Sciences,* Cambridge, Mass., MIT Press.

Gerring, John (2007) *Case study research : principles and practices,* Cambridge, Cambridge University Press.

Gilligan, Carol (1982) *In a different voice : psychological theory and women's development,* London, Harvard University Press.

Gillman, Howard (2001) What's Law Got to Do with It? Judicial Behavioralists Test the "Legal Model" of Judicial Decision Making. *Law & Social Inquiry,* 26**,** 465-504.

Gisle, Jon, Lind, Åge & Andenæs, Kristian (2010) *Jusleksikon (e-book through Ordnett.no),* Oslo, Kunnskapsforlaget.

Grendstad, Gunnar, Shaffer, William R. & Waltenburg, Eric N. (2011a) Judicial behavior on the Norwegian Supreme Court, 1945-2009. Determining the effects of extra-legal forces on the justices' votes: The case of 'government friendly' voting. *Paper prepared for presentation at 2nd International Conference on Democracy as Idea and Practice. University of Oslo, January 13-14, 2011. Workshop 4: Constitutional Democracy.* Oslo/Bergen, University of Bergen / Purdue University.

Grendstad, Gunnar, Shaffer, William R. & Waltenburg, Eric N. (2011b) When justices disagree. The influence of ideology and geography on economic voting on the Norwegian Supreme Court. *Retfærd,* 34**,** 3-22.

Grendstad, Gunnar, Shaffer, William R. & Waltenburg, Eric N. (2013a) Dommeratferd i Norges Høyesterett (Judicial Behaviour on the Supreme Court of Norway). In University, University of Bergen and Purdue (Ed.) Bergen, Norway and West Lafayette, Indiana (US).

Grendstad, Gunnar, Shaffer, William R. & Waltenburg, Eric N. (2013b) Policy Making in an Independent Judiciary. The Norwegian Supreme Court. [unpublished manuscript].

Grendstad, Gunnar; Shaffer, William R.; Waltenburg, Eric N. (2012) Public Economic Interest vs. Private Economic Rights: Preferential Voting on the Norwegian Supreme Court 1948-2011. *The 70th Annual Meeting of the Midwest Political Science Association.* Chicago, IL, USA, University of Bergen, Norway and Purdue University, IN, USA.

Grønmo, Sigmund (2004) *Samfunnsvitenskapelige metoder,* Bergen, Fagbokforlaget.

Hanssen, Inge D. (2010) Juridisk skjønn og kjønn. *Aftenposten.* Oslo, Schibsted/Aftenposten.

Hirsch, Agnes Aaby (ed.) (2010) *Dette er Kari & Ola : kvinner og menn i Norge,* Oslo, Statistisk sentralbyrå.

Hox, Joop J. (2010) *Multilevel analysis : techniques and applications,* New York, Routledge.

Jacobsen, Terje Kolbu (2012) A State Friendly Court - An Analysis of Norwegian Supreme Court Decisions. *Department of Comparative Politics.* University of Bergen.

King, Gary, Verba, Sidney & Keohane, Robert O. (1994) *Designing social inquiry : scientific inference in qualitative research,* Princeton, N.J., Princeton University Press.

Kjønstad, Asbjørn (1999) Er Høyesterett statsvennlig? *Lov og Rett***,** 97-122.

Leckie, George (2010a) *Module 5: Introduction to Multilevel Modelling (Stata Practical)* Bristol, LEMMA VLE, University of Bristol, Centre for Multilevel Modelling. Available from www.cmm.bris.ac.uk/lemma/. [Accessed on February 2nd, 2013. Cached copy available on demand if depublished.].

Leckie, George (2010b) *Module 6: Regression Models for Binary Responses (Stata Practical)* Bristol, LEMMA VLE, University of Bristol, Centre for Multilevel Modelling.

Available from www.cmm.bris.ac.uk/lemma/. [Accessed on February 2nd, 2013. Cached copy available on demand if depublished.].

Leckie, George (2013) *Module 12: Cross-Classified Multilevel Models (Stata Practical)* Bristol, LEMMA VLE, University of Bristol, Centre for Multilevel Modelling. Available from www.cmm.bris.ac.uk/lemma/. [Accessed on February 2nd, 2013. Cached copy available on demand if depublished.].

Lund, Ketil (1987) Kontroll av staten i statens egne domstoler. *Lov og Rett***,** 211-227.

Martin, Elaine & Pyle, Barry (2000) Gender, Race, and Partisanship on the Michigan Supreme Court. *Albany Law Review,* 63**,** 1205-1236.

Massie, Tajuana, Johnson, Susan W & Gubala, Sara Margaret (2002) The Impact of Gender and Race in the Decisions of Judges on the United States Courts of Appeals. *The 2002 Midwest Political Science Association Annual Meeting.* Chicago, Illinois.

McCall, Madhavi & McCall, Michael A. (2007) How far does the gender gap extend?: Decision making on state Supreme Courts in Fourth Amendment Cases, 1980–2000. *The Social Science Journal,* 44**,** 67-82.

Menkel-Meadow, Carrie (1989) Exploring a Research Agenda of the Feminization of the Legal Profession: Theories of Gender and Social Change. *Law & Social Inquiry,* 14**,** 289-319.

Midtbø, Tor (2007) *Regresjonsanalyse for samfunnsvitere,* Oslo, Universitetsforlaget.

Ministry of Children and Equality (2008) *Med barnet i fokus - En gjennomgang av barnelovens regler om foreldreansvar, bosted og samvær.* NOU 2008:9. Oslo, Departementenes Servicesenter.

Ostberg, C. L. & Wetstein, Matthew E. (2007) In a Different Voice: Sex Difference in Economic Cases Decided by the Canadian Supreme Court. *The annual meeting of the Canadian Political Science Association.* Saskatoon, Saskathewan. Canada.

Palmer, Barbara (2001) Women in the American Judiciary: Their Influence and Impact. *Women & Politics,* 23**,** 91-101.

Peresie, Jennifer L. (2005) Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts. *Yale Law Journal,* 114**,** 1759-90.

Pitkin, Hanna Fenichel (1967) *The concept of representation,* Berkeley, Calif., University of California.

Przeworski, Adam & Teune, Henry (1970) The Logic of Comparative Social Inquiry. In Przeworski, Adam & Teune, Henry (Eds.) *Comparative Research and Social Science Theory.* Malabar, Florida, Krieger Publishing Company.

Rabe-Hesketh, Sophia & Skrondal, Anders (2008) *Multilevel and Longitudinal Modeling Using Stata,* College Station, Texas, Stata Press.

Ragin, Charles (2004) Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research". In Brady, Henry E. & Collier, David (Eds.) *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* New York, Rowman and Littlefield.

Ryssdal, Anders (2006) Rettssikkerhet i forvaltningen – hvilke muligheter har du til å vinne frem. *Årstale i Den Norske Advokatforening.*

Segal, Jeffrey A. & Spaeth, Harold J. (2002) *The Supreme Court and the Attitudinal Model Revisited,* Cambridge, Cambridge University Press.

Sherry, Suzanna (1986) Civic Virtue and the Feminine Voice in Constitutional Adjudication. *Virginia Law Review,* 72**,** 543-616.

Skiple, Jon Kåre (2012) «Høyesterett er først og fremst en kollegial domstol» - Fleirnivåanalyse av effekten av ideologi på individ- og saksnivå på dommarar sine vota i økonomiske saker i Norge sin Høgsterett i tidsperioden 1991-2011. *Department of Comparative Politics.* University of Bergen.

Skivenes, Marit (2010) Judging the Child's Best Interests. *Acta Sociologica,* 53**,** 339-353.

Skjørten, Kristin (2005) *Samlivsbrudd og barnefordeling : en studie av lagmannsrettsdommer*, Oslo, Gyldendal akademisk.
Skog, Ole-Jørgen (2009) *Å forklare sosiale problemer*, Oslo, Universitetsforlaget.
Skoghøy, Jens Edvin A. (2011) Dommeratferd og dommerbakgrunn. Særlig om yrkesbakgrunnens betydning for utfallet av tvister mellom private og det offentlige. In Lambertz, Göran, Lindskog, Stefan & Möller, Mikael (Eds.) *Festskrift till Torgny Håstad*. Uppsala, Lustus.
Skåre, Jan (1997) Høyesteretts sammensetning. *Lov og Rett***,** 499-504.
Skåre, Jan & Berg Hansen, Siri (1999) Er Høyesterett statsvennlig? *Lov og Rett***,** 67-77.
StataCorp (2011) Stata Base Reference Manual: Release 12. Statistical Software. In StataCorp (Ed.) *Stata*. College Station, TX, StataCorp LP.
Steele, Fiona (2008) *Module 5: Introduction to Multilevel Modelling (Concepts)* Bristol, LEMMA VLE, University of Bristol, Centre for Multilevel Modelling Available from www.cmm.bris.ac.uk/lemma/. [Accessed on February 2nd, 2013. Cached copy available on demand if depublished.].
Steele, Fiona (2009a) *Module 6: Regression Models for Binary Responses (Concepts)* Bristol, LEMMA VLE, University of Bristol, Centre for Multilevel Modelling Available from www.cmm.bris.ac.uk/lemma/. [Accessed on March 17th, 2013. Cached copy available on demand if depublished.].
Steele, Fiona (2009b) *Module 7: Multilevel Models for Binary Responses (Concepts)* Bristol, LEMMA VLE, University of Bristol, Centre for Multilevel Modelling. Available from www.cmm.bris.ac.uk/lemma/. [Accessed on March 17th, 2013. Cached copy available on demand if depublished.].
Steffensmeier, Darrell & Herbert, Chris (1999) Women and Men Policymakers: Does the Judge's Gender Affect the Sentencing of Criminal Defendants? *Social Forces,* 77**,** 1163-1196.
Sullivan, Kathleen M. (2002) Constitutionalising Women's Equality. *California Law Review,* 90**,** 735-64.
Sunde, Jørn Øyrehagen (2012a) Debatt: Andre premiss og anna resultat - refleksjonar kring politikk, Høgsterett og dissensar. *Tidsskrift for rettsvitenskap,* 125**,** 168-204.
Sunde, Jørn Øyrehagen (2012b) Dissenting votes in the Norwegian Supreme Court 1965-2009: A legal cultural analysis. *Rechtskultur - Zeitschrift für Europäische Rechtsgeschichte***,** 59-73.
Tellesbø, Ola (2006) Hvorfor Høyesterett er statsvennlig. *Rettfærd,* 29**,** 65-78.
Vaage, Odd Frank (2012) *Tidene skifter : tidsbruk 1971-2010,* Oslo-Kongsvinger, Statistisk sentralbyrå.
Wold, Terje (1964) Domstolenes deltakelse i justisforvaltningen. *Lov og Rett***,** 385.
Østlid, Henry (1988) *Dommeratferd i dissenssaker,* Oslo, Universitetsforlaget.

Appendix

7.1 Cross-classification. Stata output.

```
. table JNR DomNr if inrange(JNR, 101, 130) & inrange(DomNr,3306, 5482)
```

	DomNr																
JNR	3306	4703	4734	4742	4789	4815	4890	5013	5014	5084	5146	5148	5224	5339	5383	5478	5482
101	1																
102				1		1						1		1			
103		1	1						1	1		1	1	1	1	1	1
104				1	1						1						
105	1					1		1	1				1				1
106						1	1			1		1					
107	1		1					1	1	1	1			1		1	1
109														1			
110					1	1				1			1				
111						1					1	1					
112							1								1		
113	1												1			1	
114		1	1		1		1		1								
115		1						1							1		
116				1						1		1		1			1
117								1	1								
118							1										
119			1	1													
120		1			1												
121		1															
125							1	1							1		1
126											1		1				
127											1						
129																1	

Note: This is an illustrative excerpt.

7.2 Frequency of observations per judge appointment

NamnNummerKvinne	Freq.	Percent	Cum.
Arne Christiansen, 355, 0	34	4.53	4.53
Gunnar Aasland, 125, 0	33	4.40	8.93
Vera Louise Holmøy, 130, 1	30	4.00	12.93
Tore Sinding-Larsen, 311, 0	29	3.87	16.80
Liv Gjølstad, 103, 1	28	3.73	20.53
Rolv Hellesylt, 218, 0	25	3.33	23.87
Elisabeth Schweigaard Selmer, 307, 1	23	3.07	26.93
Jan Rasmus Skåre, 316, 0	21	2.80	29.73
Jens Bugge, 133, 0	21	2.80	32.53
Knut Blom, 159, 0	21	2.80	35.33
Steinar Tjomsland, 106, 0	20	2.67	38.00
Karenanne Gussgard, 105, 1	19	2.53	40.53
Helge Røstad, 295, 0	18	2.40	42.93
Lilly Helena Bølviken, 170, 1	17	2.27	45.20
Kirsti Coward, 107, 1	16	2.13	47.33
Finn Backer, 356, 0	14	1.87	49.20
Hans Methlie Michelsen, 262, 0	14	1.87	51.07
Trond Dolva, 127, 0	14	1.87	52.93
Andreas Endresen, 182, 0	13	1.73	54.67
Egil Endresen, 183, 0	13	1.73	56.40
Jan Frøystein Halvorsen, 206, 0	13	1.73	58.13
Ketil Lund, 104, 0	13	1.73	59.87
Nils Peder Langvand, 245, 0	13	1.73	61.60
Tore Schei, 101, 0	13	1.73	63.33
Einar Løchen, 354, 0	11	1.47	64.80
Hans Flock, 110, 0	11	1.47	66.27
Carl Ludovico Stabel, 320, 0	10	1.33	67.60
Charles Philipson, 276, 0	10	1.33	68.93
Jens Christian Mellbye, 260, 0	10	1.33	70.27
Magnus Aarbakke, 131, 0	10	1.33	71.60
Sigurd Juell Lorentzen, 253, 0	10	1.33	72.93
Trygve Leivestad, 249, 0	10	1.33	74.27
Henrik Eiler Støren Bahr, 382, 0	9	1.20	75.47
Jens Edvin A. Skoghøy, 114, 0	9	1.20	76.67
Rolv Einar Ryssdal, 294, 0	9	1.20	77.87
Atle Roll-Matthiesen, 290, 0	8	1.07	78.93
Finn Hiorthøy, 221, 0	8	1.07	80.00
Johan Eilert Stang Lund, 323, 0	8	1.07	81.07
Karin Maria Bruzelius, 113, 1	8	1.07	82.13
Per Tønseth, 338, 0	8	1.07	83.20
Axel Heiberg, 216, 0	7	0.93	84.13
Harald Magne Elstad, 181, 0	7	0.93	85.07
Carsten Smith, 129, 0	6	0.80	85.87
Ernst Fredrik Eckhoff, 178, 0	6	0.80	86.67
Ingse Stabel, 116, 1	6	0.80	87.47
Oscar Christian Gundersen, 203, 0	6	0.80	88.27
Finn Backer, 431, 0	5	0.67	88.93
Karsten Gaarder, 204, 0	5	0.67	89.60
Tore Schei, 102, 0	5	0.67	90.27
Trygve Bendiksby, 152, 0	5	0.67	90.93
Erling Sandene, 298, 0	4	0.53	91.47
Erling Sandene, 299, 0	4	0.53	92.00
Karl Arne Utgård, 115, 0	4	0.53	92.53
Lars Oftedal Broch, 109, 0	4	0.53	93.07
Magnus Matningsdal, 111, 0	4	0.53	93.60
Bård Tønder, 119, 0	3	0.40	94.00
Erik Arnt Foss, 188, 0	3	0.40	94.40
Georg Fredrik Rieber-Mohn, 112, 0	3	0.40	94.80
Marius Nygaard, 269, 0	3	0.40	95.20
Clement Endresen, 120, 0	2	0.27	95.47
Federik Zimmer, 446, 0	2	0.27	95.73
Hilde Indreberg, 121, 1	2	0.27	96.00
Ole Bjørn Støle, 117, 0	2	0.27	96.27
Rolv Einar Ryssdal, 293, 0	2	0.27	96.53
Sverre Mitsem, 126, 0	2	0.27	96.80
Toril Marie Øie, 118, 1	2	0.27	97.07
Arne Christiansen, 453, 0	1	0.13	97.20
Astri Sverdrup Rynning, 471, 1	1	0.13	97.33
Bergljot Webster, 359, 1	1	0.13	97.47
Birger Stuevold Lassen, 437, 0	1	0.13	97.60
Birger Stuevold Lassen, 438, 0	1	0.13	97.73
Christian Borchsenius, 466, 0	1	0.13	97.87
Erling Haugen, 212, 0	1	0.13	98.00
Gustav M. Sverdrup-Thygeson, 335, 0	1	0.13	98.13
Jens Fagereng, 185, 0	1	0.13	98.27
Jørgen Berner Thrap, 333, 0	1	0.13	98.40
Knut Herbrand Kallerud, 462, 0	1	0.13	98.53
Kristen Syvertsen, 426, 0	1	0.13	98.67
Kristin Normann, 363, 1	1	0.13	98.80
Magnus Aarbakke, 469, 0	1	0.13	98.93
Nina Frisak, 132, 1	1	0.13	99.07
Otto Helgesen, 217, 0	1	0.13	99.20
Per Lykke Anker, 141, 0	1	0.13	99.33
Peter Lødrup, 258, 0	1	0.13	99.47
Peter Lødrup, 441, 0	1	0.13	99.60
Peter Lødrup, 444, 0	1	0.13	99.73
Reidar Dick Henriksen, 424, 0	1	0.13	99.87
Sverre Dragsten, 176, 0	1	0.13	100.00
Total	750	100.00	

Note: This table lists the frequency of observations per judge appointment, not per judge. Each judge is listed for every time they have been appointed, e.g. Peter Lødrup is registered with one child custody case for each of his three appointment periods (almost at the bottom of the table). Similarly, Tore Schei is appointed two times, the first time as Supreme Court judge, and the second time as chief justice ("justitiarius") of the Supreme Court. As can be observed in the table on the next page, these two sets of observations are added together when counting how many observations a judge has partaken in. Rather than either 5 or 13 observations, Schei is thus listed with 18 observations in the following table, which is the correct number.

7.3 Identifying critical actors

Table: Identifying critical actors - the full table

Judge		**Bivariate – consistent voting**				**Trivariate – peer effect**			
(frequency in selection)		coeff.	s.e.	p-value	-2LL†	coeff.	s.e.	p-value	-2LL†
Leivestad	*individual*	-3.69**	1.84	.046	**	-3.90**	1.98	.049	.623
(10)	*in panel*				.016	1.17	2.39	.623	
Lorentzen	*individual*	2.33	1.42	.102	*	2.13	1.39	.126	.170
(10)	*in panel*				.068	3.01	2.26	.183	
Aarbakke	*individual*	-1.14	1.39	.410	.400	-1.11	1.39	.426	.843
(10)	*in panel*					-.44	2.22	.843	
Mellbye	*individual*	.32	1.34	.808	.809	.37	1.35	.784	.785
(10)	*in panel*					-.63	2.29	.785	
Philipson	*individual*	.68	1.84	.713	.709	.67	1.86	.720	.978
(10)	*in panel*					.06	2.39	.979	
Stabel	*individual*	-.38	2.82	.894	.893	<.00	2.89	>.99	.458
(10)	*in panel*					-2.35	3.20	.463	
Flock	*individual*	.374	1.31	.775	.776	.35	1.32	.789	.897
(11)	*in panel*					.29	2.21	.897	
Løchen	*individual*	1.41	1.62	.384	.361	1.40	1.64	.392	.979
(11)	*in panel*					.06	2.26	.979	
Schei	*individual*	-.36	1.12	.747	.400	-.24	1.12	.826	.744
(18)	*in panel*					-1.50	1.79	.402	
Langvand	*individual*	-1.50	1.60	.351	.338	-1.47	1.62	.365	.915
(13)	*in panel*					-.23	2.15	.915	
Lund	*individual*	2.45	1.72	.153	*	2.12	1.60	.185	.136
(13)	*in panel*				.088	3.02	2.06	.142	
Halvorsen	*individual*	.52	1.29	.689	.689	.69	1.33	.603	.336
(13)	*in panel*					-2.01	2.10	.341	
E. Endresen	*individual*	-1.50	1.37	.275	.249	-1.32	1.36	.332	.233
(13)	*in panel*					-2.45	2.08	.240	
A. Endresen	*individual*	-.66	.95	.485	.483	-.73	.96	.451	.532
(13)	*in panel*					1.22	1.95	.533	
Dolva	*individual*	-2.64	1.78	.138	.102	-2.42	1.74	.164	.518
(14)	*in panel*					-1.35	2.10	.520	
Michelsen	*individual*	.731	1.65	.657	.651	.52	1.63	.751	.420
(14)	*in panel*					1.65	2.06	.422	
Backer	*individual*	1.62	1.12	.149	.141	1.75	1.16	.133	.513
(14)	*in panel*					-1.17	1.79	.514	
Coward		-2.10	1.39	.131	.095	-2.08	1.40	.137	.933

(16)						-.16	1.90	.933	
Bølviken		-1.48	.96	.123	.115	-1.64*	.99	.098	.224
(17)						2.18	1.80	.226	
Røstad		.88	1.34	.510	.511	1.17	1.42	.407	.243
(18)						-2.19	1.91	.250	
Gussgard		-1.59	1.11	.152	.123	-1.43	1.10	.193	.146
(19)						-2.51	1.75	.153	
Tjomsland		-2.83**	1.29	.029	***	-2.59**	1.25	.039	*
(20)					.010	-2.88*	1.75	.099	.090
Blom		.14	.91	.878	.877	.00	.92	>.99	.167
(21)						2.26	1.66	.174	
Bugge		.50	.87	.566	.564	.45	.87	.604	.617
(21)						.82	1.64	.618	
Skåre		-.61	1.16	.597	.593	-.53	1.16	.648	.605
(21)						-.88	.71	.606	
Schw. Selmer		-2.07**	1.03	.045	.030	-2.12**	1.05	.044	.739
(23)						0.54	1.62	.739	
Hellesylt		-.30	.97	.755	.753	-.19	.97	.847	.335
(25)						-1.51	1.58	.339	
Gjølstad		-.99	1.26	.433	.418	-.60	1.24	.630	**
(28)						-3.21**	1.60	.044	.037
Sinding-Larsen		1.76*	.94	.061	**	1.64*	.93	.078	.318
(29)					.046	1.47	1.49	.322	
Holmøy		1.69**	.86	.050	**	1.51*	.85	.074	*
(30)					.038	2.41	1.47	.100	.091
Aasland		-.11	.92	.906	.906	<-.01	.93	>.99	.366
(33)						-1.29	1.43	.369	
Christiansen		.14	.83	.865	.865	.17	.83	.843	.812
(34)						-.33	1.39	.813	

*Note: The dependent variable is individual judges' propensity to vote for the feminine party. †-2LL in the bivariate models are comparisons with the empty model. Equally, the trivariate model is compared to the bivariate. The statistic is given as a p-value. The analyses are run using the xtmelogit command with 30 integration points in Stata. *** p<.01, ** p<.05, * p<.10. See also note to the table in appendix section 7.2.*

7.3.1 Justice Gjølstad

Justice Gjølstad is not identified as a critical actor in the bivariate model. Counter-intuitively, in the trivariate model the panel level variable indicates that in the panels she has partaken, her peers have been voting significantly consistent to the disadvantage of the female party. The direction of this effect is similar to Gjølstad's own (non-significant) voting pattern in this selection of cases. Gjølstad is not, however, identified as a critical actor, because the operational definition of a critical actor demands that the judge' individual voting pattern is significant on a five per cent level. For the sake of interest, an attempt at adding Gjølstad to the analysis saw the "Gjølstad effect" turn insignificant in a model including the significant variables from the steps before the critical actors analysis. Outputs of this particular analysis is not added to the appendix, but can be presented on demand.

7.3.2 All critical actors in one model

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  24                        Wald chi2(5)       =     17.41
Log likelihood =  -280.4003                     Prob > chi2        =    0.0038
```

VoteForFemi~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Leivestad	-3.761165	1.886738	-1.99	0.046	-7.459103	-.063227
Tjomsland	-2.638332	1.325319	-1.99	0.047	-5.23591	-.0407533
Selmer	-1.981709	1.050226	-1.89	0.059	-4.040114	.0766949
Holmoy	1.60954	.8838198	1.82	0.069	-.1227153	3.341795
SindingLarsen	1.651037	.9762051	1.69	0.091	-.2622894	3.564364
_cons	-.7846909	.6181144	-1.27	0.204	-1.996173	.4267911

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
var(_cons)	41.60875	12.4469	23.15022	74.78495

```
LR test vs. logistic regression: chibar2(01) =   452.35 Prob>=chibar2 = 0.0000
```

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	750	.	-280.4003	7	574.8006	607.1411

Note: N=Obs used in calculating BIC; see **[R] BIC note**

```
. estat icc

Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9267267	.020313	.8755727	.9578626

```
. lrtest Tom

Likelihood-ratio test                                 LR chi2(5)  =     23.84
(Assumption: Tom nested in .)                         Prob > chi2 =    0.0002
```

7.4 Individual block 4.1

```
Model Indblokk41
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  30                        Wald chi2(3)       =     13.86
Log likelihood = -284.53658                     Prob > chi2        =    0.0031
```

VoteForFem~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-.8220402	.4324963	-1.90	0.057	-1.669717	.025637
OsloBorn	.838048	.3409328	2.46	0.014	.1698321	1.506264
c_YGRADCJ	-.0420392	.0216116	-1.95	0.052	-.0843971	.0003187
_cons	-1.03266	.629973	-1.64	0.101	-2.267385	.202064

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
sd(_cons)	6.338149	.9717316	4.69311	8.559811

```
LR test vs. logistic regression: chibar2(01) =   450.40 Prob>=chibar2 = 0.0000

. lrtest Indblokk41 Indblokk3

Likelihood-ratio test                                 LR chi2(1)  =      3.91
(Assumption: Indblokk3 nested in Indblokk41)          Prob > chi2 =    0.0479

. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
Indblokk41	750	.	-284.5366	5	579.0732	602.1735

Note: N=Obs used in calculating BIC; see **[R] BIC note**

```
. estat icc

Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9243047	.0214535	.8700435	.957029

7.5 Individual block 5.2

```
Model Indblokk51
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  30                        Wald chi2(4)       =     15.90
Log likelihood = -283.08215                     Prob > chi2        =    0.0032
```

VoteForFem~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-.8151085	.4385422	-1.86	0.063	-1.674635	.0444184
OsloBorn	.8451549	.3465025	2.44	0.015	.1660224	1.524287
c_YGRADCJ	-.0442879	.0218006	-2.03	0.042	-.0870164	-.0015594
FormJudge	-.5694841	.3369143	-1.69	0.091	-1.229824	.0908558
_cons	-.7867589	.6481388	-1.21	0.225	-2.057088	.4835698

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
sd(_cons)	6.381785	.979539	4.723799	8.621701

```
LR test vs. logistic regression: chibar2(01) =   450.58 Prob>=chibar2 = 0.0000

. lrtest Indblokk41 Indblokk51

Likelihood-ratio test                                 LR chi2(1)  =      2.91
(Assumption: Indblokk41 nested in Indblokk51)         Prob > chi2 =    0.0881

. estat icc

Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9252592	.0212291	.8715103	.9576177

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
Indblokk51	750	.	-283.0821	6	578.1643	605.8847

```
Note:  N=Obs used in calculating BIC; see [R] BIC note
```

7.6 Panel level block 3.1

Model Kollblokk31

Mixed-effects logistic regression — Number of obs = 750
Group variable: DomNr — Number of groups = 150

Obs per group: min = 5
avg = 5.0
max = 5

Integration points = 24 — Wald chi2(4) = 25.28
Log likelihood = -275.47937 — Prob > chi2 = 0.0000

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-.8595779	.426734	-2.01	0.044	-1.695961	-.0231946
OsloBorn	.855175	.339838	2.52	0.012	.1891048	1.521245
c_YGRADCJ	-.0215477	.0215833	-1.00	0.318	-.0638501	.0207547
OpponentPublic	-5.164881	1.325916	-3.90	0.000	-7.763629	-2.566133
_cons	.7404696	.6988971	1.06	0.289	-.6293435	2.110283

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
sd(_cons)	5.745003	.8195544	4.343721	7.598337

LR test vs. logistic regression: chibar2(01) = 397.18 Prob>=chibar2 = 0.0000

```
. lrtest Kollblokk31 IndEndeleg
```

Likelihood-ratio test — LR chi2(1) = 18.11
(Assumption: IndEndeleg nested in Kollblokk31) — Prob > chi2 = 0.0000

```
. lrtest Kollblokk31 Kollblokk3
```

Likelihood-ratio test — LR chi2(2) = 0.32
(Assumption: Kollblokk31 nested in Kollblokk3) — Prob > chi2 = 0.8517

```
. estat icc
```

Residual intraclass correlation

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9058255	.0239698	.8472215	.9434501

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
Kollblokk32	750	.	-275.9876	5	561.9753	585.0756

Note: N=Obs used in calculating BIC; see **[R] BIC note**

7.7 Panel level block 3.2

```
Model Kollblokk32
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  24                        Wald chi2(3)       =     25.35
Log likelihood = -275.98764                     Prob > chi2        =    0.0000
```

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-.9099165	.4205004	-2.16	0.030	-1.734082	-.0857508
OsloBorn	.877114	.3373312	2.60	0.009	.215957	1.538271
OpponentPublic	-5.337178	1.294662	-4.12	0.000	-7.874669	-2.799688
_cons	.8233884	.681637	1.21	0.227	-.5125956	2.159372

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
sd(_cons)	5.625288	.7903157	4.271266	7.408546

```
LR test vs. logistic regression: chibar2(01) =   398.25 Prob>=chibar2 = 0.0000

. lrtest Kollblokk32 Kollblokk3

Likelihood-ratio test                                 LR chi2(3)  =      1.34
(Assumption: Kollblokk32 nested in Kollblokk3)        Prob > chi2 =    0.7202

. estat icc

Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9058255	.0239698	.8472215	.9434501

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
Kollblokk32	750	.	-275.9876	5	561.9753	585.0756

Note: N=Obs used in calculating BIC; see **[R] BIC note**

7.8 Critical actors: gradual removal of variables

7.8.1 First: removing Schweigaard Selmer

```
Model KritAkt11
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  24                        Wald chi2(7)       =     35.82
Log likelihood = -264.82982                     Prob > chi2        =    0.0000
```

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-1.600196	.5206363	-3.07	0.002	-2.620625	-.5797678
OsloBorn	.5946888	.3700231	1.61	0.108	-.1305432	1.319921
OpponentPublic	-5.63004	1.396446	-4.03	0.000	-8.367023	-2.893057
Leivestad	-4.085986	2.00167	-2.04	0.041	-8.009187	-.1627852
Tjomsland	-2.405556	1.272608	-1.89	0.059	-4.899822	.0887111
SindingLarsen	1.748753	1.018185	1.72	0.086	-.2468536	3.744359
Holmoy	2.615655	1.026348	2.55	0.011	.6040502	4.62726
_cons	1.084701	.7449848	1.46	0.145	-.3754423	2.544845

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
var(_cons)	36.71794	10.6708	20.77333	64.90088

```
LR test vs. logistic regression: chibar2(01) =   406.69 Prob>=chibar2 = 0.0000

. estat icc

Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9177693	.0219324	.8632822	.9517549

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	750	.	-264.8298	9	547.6596	589.2403

```
Note:  N=Obs used in calculating BIC; see [R] BIC note

. lrtest KritAkt11 KritAkt1

Likelihood-ratio test                                 LR chi2(1)  =      1.72
(Assumption: KritAkt11 nested in KritAkt1)            Prob > chi2 =    0.1896
```

7.8.2 Second: removing the geographical variable

```
Model KritAkt12
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  24                        Wald chi2(6)       =     34.26
Log likelihood = -266.13724                     Prob > chi2        =    0.0000
```

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-1.688566	.5169975	-3.27	0.001	-2.701862	-.6752694
OpponentPublic	-5.60689	1.390891	-4.03	0.000	-8.332986	-2.880793
Leivestad	-4.280411	1.995509	-2.15	0.032	-8.191536	-.369286
Tjomsland	-2.610799	1.265762	-2.06	0.039	-5.091647	-.1299518
SindingLarsen	1.474396	.9969353	1.48	0.139	-.4795611	3.428353
Holmoy	3.034191	.9967032	3.04	0.002	1.080689	4.987694
_cons	1.339123	.7276914	1.84	0.066	-.0871259	2.765372

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
var(_cons)	36.42564	10.57521	20.61976	64.34738

```
LR test vs. logistic regression: chibar2(01) =   407.02 Prob>=chibar2 = 0.0000

. estat icc

Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9177693	.0219324	.8632822	.9517549

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	750	.	-264.8298	9	547.6596	589.2403

```
Note:  N=Obs used in calculating BIC; see [R] BIC note

. lrtest KritAkt12 KritAkt1

Likelihood-ratio test                                 LR chi2(2)  =      4.34
(Assumption: KritAkt12 nested in KritAkt1)            Prob > chi2 =    0.1144

. lrtest KritAkt12 KritAkt11

Likelihood-ratio test                                 LR chi2(1)  =      2.61
(Assumption: KritAkt12 nested in KritAkt11)           Prob > chi2 =    0.1059
```

7.8.3 Third: removing justice Sinding-Larsen

```
Model KritAkt13
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  24                        Wald chi2(5)       =     33.29
Log likelihood = -267.33477                     Prob > chi2        =    0.0000
```

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-1.765401	.514283	-3.43	0.001	-2.773377	-.7574249
OpponentPublic	-5.551661	1.379305	-4.02	0.000	-8.255049	-2.848274
Leivestad	-4.296769	1.994673	-2.15	0.031	-8.206256	-.3872807
Tjomsland	-2.737883	1.260377	-2.17	0.030	-5.208175	-.2675902
Holmoy	3.039316	.9930617	3.06	0.002	1.092951	4.985681
_cons	1.399448	.7231134	1.94	0.053	-.0178279	2.816724

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
var(_cons)	36.00468	10.40239	20.43769	63.42873

```
LR test vs. logistic regression: chibar2(01) =   408.05 Prob>=chibar2 = 0.0000

. estat icc

Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9177693	.0219324	.8632822	.9517549

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	750	.	-264.8298	9	547.6596	589.2403

```
Note:  N=Obs used in calculating BIC; see [R] BIC note

. lrtest KritAkt13 KritAkt1

Likelihood-ratio test                                 LR chi2(3)  =      6.73
(Assumption: KritAkt13 nested in KritAkt1)            Prob > chi2 =    0.0810

. lrtest KritAkt13 KritAkt11

Likelihood-ratio test                                 LR chi2(2)  =      5.01
(Assumption: KritAkt13 nested in KritAkt11)           Prob > chi2 =    0.0817
```

7.8.4 Fourth: testing a model without Tjomsland

```
Model KritAkt14
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  24                        Wald chi2(4)       =     29.92
Log likelihood = -270.43787                     Prob > chi2        =    0.0000
```

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-1.677458	.5089478	-3.30	0.001	-2.674978	-.6799391
OpponentPublic	-5.598784	1.357781	-4.12	0.000	-8.259986	-2.937583
Leivestad	-4.219383	1.961388	-2.15	0.031	-8.063633	-.375133
Holmoy	3.076648	.9850448	3.12	0.002	1.145995	5.0073
_cons	1.34034	.7085682	1.89	0.059	-.0484284	2.729108

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Identity				
var(_cons)	34.59947	9.912335	19.73378	60.66364

```
LR test vs. logistic regression: chibar2(01) =   404.65 Prob>=chibar2 = 0.0000

. estat icc

Residual intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
DomNr	.9177693	.0219324	.8632822	.9517549

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	750	.	-264.8298	9	547.6596	589.2403

```
Note:  N=Obs used in calculating BIC; see [R] BIC note

. lrtest KritAkt14 KritAkt1

Likelihood-ratio test                                 LR chi2(4)  =     12.94
(Assumption: KritAkt14 nested in KritAkt1)            Prob > chi2 =    0.0116

. lrtest KritAkt14 KritAkt11

Likelihood-ratio test                                 LR chi2(3)  =     11.22
(Assumption: KritAkt14 nested in KritAkt11)           Prob > chi2 =    0.0106
```

7.9 Testing for varying slopes

7.9.1 Varying slope: gender

Model RSsex

Mixed-effects logistic regression — Number of obs = 750
Group variable: DomNr — Number of groups = 150

Obs per group: min = 5
avg = 5.0
max = 5

Integration points = 24 — Wald chi2(5) = 28.50
Log likelihood = -267.14296 — Prob > chi2 = 0.0000

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-1.692499	.5980194	-2.83	0.005	-2.864596	-.5204028
OpponentPublic	-5.679116	1.418795	-4.00	0.000	-8.459904	-2.898329
Leivestad	-4.3693	2.05611	-2.13	0.034	-8.399201	-.3393992
Tjomsland	-2.782206	1.277825	-2.18	0.029	-5.286697	-.2777139
Holmoy	2.979628	1.149635	2.59	0.010	.7263845	5.232871
_cons	1.435956	.7451064	1.93	0.054	-.0244258	2.896338

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Unstructured				
sd(SEX)	.8049932	1.283331	.0353831	18.31425
sd(_cons)	6.187708	.961168	4.563609	8.389793
corr(SEX,_cons)	-.5714386	1.534522	-.9999279	.9990316

LR test vs. logistic regression: chi2(3) = 408.43 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

```
. lrtest RSsex KritAkt3
```

Likelihood-ratio test — LR chi2(2) = 0.38
(Assumption: KritAkt3 nested in RSsex) — Prob > chi2 = 0.8255

7.9.2 Varying slope: justice Leivestad

```
Model RSleiv
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  24                        Wald chi2(5)       =     33.29
Log likelihood = -267.33477                     Prob > chi2        =    0.0000
```

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-1.765401	.514283	-3.43	0.001	-2.773377	-.757425
OpponentPublic	-5.551662	1.379305	-4.02	0.000	-8.255049	-2.848274
Leivestad	-4.29677	1.994674	-2.15	0.031	-8.206259	-.3872818
Tjomsland	-2.737882	1.260376	-2.17	0.030	-5.208175	-.2675898
Holmoy	3.039317	.9930618	3.06	0.002	1.092951	4.985682
_cons	1.399449	.7231133	1.94	0.053	-.0178275	2.816724

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Independent				
sd(Leives~d)	2.81e-09	4.068471	0	.
sd(_cons)	6.00039	.866809	4.520807	7.964216

```
LR test vs. logistic regression:     chi2(2) =   408.05   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. lrtest RSleiv KritAkt3

Likelihood-ratio test                                 LR chi2(1)  =     -0.00
(Assumption: KritAkt3 nested in RSleiv)               Prob > chi2 =    1.0000
```

7.9.3 Varying slope: justice Tjomsland

Model RStjom

Mixed-effects logistic regression — Number of obs = 750
Group variable: DomNr — Number of groups = 150

Obs per group: min = 5
avg = 5.0
max = 5

Integration points = 24 — Wald chi2(5) = 30.69
Log likelihood = -267.33217 — Prob > chi2 = 0.0000

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-1.767587	.5156193	-3.43	0.001	-2.778182	-.7569919
OpponentPublic	-5.555392	1.381839	-4.02	0.000	-8.263746	-2.847038
Leivestad	-4.300346	1.997119	-2.15	0.031	-8.214627	-.3860647
Tjomsland	-2.818157	1.747187	-1.61	0.107	-6.24258	.6062658
Holmoy	3.045216	.9966453	3.06	0.002	1.091827	4.998605
_cons	1.401208	.7244073	1.93	0.053	-.0186037	2.821021

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Independent				
sd(Tjomsl~d)	.7062069	5.212163	3.69e-07	1352784
sd(_cons)	6.007496	.8743487	4.516553	7.990609

LR test vs. logistic regression: chi2(2) = 408.05 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

```
. lrtest RStjom KritAkt3
```

Likelihood-ratio test — LR chi2(1) = 0.01
(Assumption: KritAkt3 nested in RStjom) — Prob > chi2 = 0.9425

7.9.4 Varying slope: justice Holmøy

```
Model RSholm
```

```
Mixed-effects logistic regression               Number of obs      =       750
Group variable: DomNr                           Number of groups   =       150

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

Integration points =  24                        Wald chi2(5)       =     32.01
Log likelihood =  -267.2862                     Prob > chi2        =    0.0000
```

VoteForFemin~e	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SEX	-1.778039	.5179159	-3.43	0.001	-2.793136	-.7629427
OpponentPublic	-5.576088	1.393265	-4.00	0.000	-8.306838	-2.845339
Leivestad	-4.321762	2.008295	-2.15	0.031	-8.257949	-.3855754
Tjomsland	-2.751529	1.264999	-2.18	0.030	-5.230882	-.2721764
Holmoy	3.151903	1.135263	2.78	0.005	.9268291	5.376978
_cons	1.401246	.7291664	1.92	0.055	-.0278941	2.830386

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
DomNr: Independent				
sd(Holmoy)	1.158954	2.138947	.0311245	43.1549
sd(_cons)	6.05271	.8936833	4.531791	8.084066

```
LR test vs. logistic regression:     chi2(2) =   408.15   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. lrtest RSholm KritAkt3

Likelihood-ratio test                                 LR chi2(1)  =      0.10
(Assumption: KritAkt3 nested in RSholm)               Prob > chi2 =    0.7553
```