Decision-Making in the European Court of Human Rights: Quantifying the European Consensus Method

Hauke Licht[1], Julia Maynard[2], Maria Fanou[3], Nicole Baerg[4], Panos Kapotas[5], Kushtrim Istrefi[6], Jonathan Slapin[7], Denise Traber[8], Vassilis Tzevelekos[9]

31 August 2022

Word Count: Approx. 10,800

Abstract

When deciding cases about the emergence of human rights across Europe, Judges at the European Court of Human Rights (ECtHR) use methods of interpretation to provide reasoning or justification for their decisions. The nature and quality of this legal argumentation is important because it can provide the Court's ruling with legitimacy, especially within the national legal systems of the European Convention of Human Rights (ECHR) contracting states. In this paper, we explore the nature of one of these legal methods of interpretation - European consensus (EuC). The Court often uses EuC when making decisions on morally, politically, or socially controversial, sensitive, or ambiguous human rights (HR) issues. But as a method of legal interpretation, EuC remains incompletely defined and understood by the legal community. Scholars can be unsure of what constitutes consensus and, in many instances -in particular when this term is not explicitly used by a ECtHR judgment - even whether the Court has employed it as a method of interpretation. We apply quantitative text-as-data methods to quantify the use of EuC within ECtHR decision-making. We reveal the language used in conjunction with the method and we demonstrate the frequency with which the Court uses the method, how this has changed over time, and the correlates of EuC, such as the nature of cases where it is used by the court.

[1] University of Cologne
[2] University of Michigan, University of Zurich
[3] University of Zurich
[4] University of Essex
[5] University of Portsmouth
[6] Utrecht University
[7] University of Zurich
[8] University of Basel
[9] University of Liverpool

Introduction

Judges use methods of legal interpretation, or legal "tools", to construct arguments when deciding cases. This is not only true of judges in national legal systems, but also of international judges. Judges on the European Court of Human Rights (ECtHR) -- the international court that adjudicates cases regarding human rights outlined in the European Convention of Human Rights (ECHR) -- have developed a particular method of legal interpretation, the method of European consensus (EuC). EuC is a "tool" used by ECtHR judges when making potentially controversial decisions about the emergence of new human rights, e.g., the right to have an abortion, or the right to same-sex marriage. It is a method of comparison to other countries' laws and norms, but also to other sources of law, such as international law standards, the case law of other (international) courts, or the practice of international organizations. In short, EuC allows the Court to draw on external sources and human rights standards as a means of argumentation. Judges can, in effect, argue that it is the consensus of a variety of sources that a new right has or has not emerged. Such arguments, as asserted by scholars of European human rights law and the ECtHR, can confer legitimacy on ECtHR decisions and increase the likelihood that actors (e.g., judges and lawyers) in national legal systems may pay them heed (Dzehtsiarou 2015).

Despite its importance in the development of European human rights law, EuC remains understudied and incompletely understood. Legal scholars with an interest in understanding legal argumentation and its impacts on European human rights law have not fully accounted for the use of EuC across the entirety of ECtHR case law. They tend to focus instead on the most high-profile and known instances of its use, without knowing, or investigating, whether these instances are representative of the Court's use of EuC, more generally. In contrast to legal scholars, social scientists interested in international courts, international human rights law, and

the ECtHR, in particular, have engaged in "large-N" studies of ECtHR decision-making, but have focused their attention on the nature of judicial outcomes, state compliance with those outcomes, and public attitudes towards the ECtHR, rather than the nature of the Court's legal argumentation. They may have not examined the nature of legal argumentation perhaps because the nature of legal arguments is often less salient or important in the view of scholars outside the field of law.

In this project, we apply established quantitative text analysis methods, which allows us to better understand the process of legal decision-making and nature of legal argumentation, looking beyond case outcomes and their impact on international human rights law (see e.g. Voeten 2007, 2008, 2021). Until now, text-as-data methods as applied to the decisions of the ECtHR have been primarily used to predict possible outcomes of cases (Aletras et al. 2016; Medvedeva, Vols, and Wieling 2020), but they have not been used to understand the methods of legal interpretation that the Court actually uses to reach its decisions (Peat 2021, Theilen 2021).

Our goal in this paper is to uncover the evolution of the ECtHR's use of the EuC method in its decision-making. We do so by focusing on the Court's Grand Chamber judgments. These cases are the most important and consequential of cases adjudicated by the ECtHR and they are heard by 17 judges, including the Court's President and Vice-Presidents. Cases reach the Grand Chamber either after judges in a Chamber proceeding (consisting of 7 judges) have issued a ruling and one of the parties involved requests a referral to the Grand Chamber or if the Chamber, because of the gravity or importance of a case, relinquishes jurisdiction to the Grand Chamber. Thus, by focusing on Grand Chamber judgments, we cover the most important decisions made by the ECtHR.

Identifying and mapping the use of the EuC method over time in the ECtHR Grand Chamber will lead us to a better understanding of how human rights law has changed and developed across Europe. It also provides the basis for answering more theoretically driven questions, such as when and why the ECtHR uses EuC when making judgments. Is EuC more prevalent in certain types of cases, amongst certain judges, or when certain ECHR parties are involved? After introducing the EuC method and discussing how ECtHR and judicial decision-making has been studied to date, we take a text-as-data approach to identify the presence of EuC reasoning in judicial decision-making. We hand-code a significant number of ECtHR Grand Chamber judgments according to rigorous coding procedure and then train a classifier to identify the nature of EuC within the ECtHR's decisions and how the use of this legal tool has changed and evolved over time. Doing so helps us to understand the formal and informal decision-making tools and modes of legal argumentation that ECtHR Judges use when deciding about the nature of human rights in Europe. We study when and where EuC is used to gain an understanding not only of how ECtHR judges make reasoned arguments and reach decisions but also to understand how human rights standards emerge across Europe. We find that the use of EuC as a tool for legal decision-making has increased significantly since the inception of the Court, and we discuss the challenges of identifying EuC language in the Grand Chamber.

What is European Consensus?

EuC is a "tool" of legal interpretation invented by and often used by the ECtHR when making decisions on morally, politically, or socially controversial, sensitive, or ambiguous human rights issues, such as abortion, same-sex marriage, or religious clothing. In a nutshell, EuC is a comparative approach in which judges compare practices in other states and

organizations to trace the evolution of societal norms regarding human rights. Using EuC enables adjudicators to interpret human rights law in a dynamic manner that reflects present-day conditions and attitudes. This, in turn, opens opportunities for change of norms via interpretation.

Narrowly conceived, EuC consists of the comparative analysis of the laws and practices of the Council of Europe (CoE) member states on the regulation of the human rights question at issue, with a view to identifying whether a new shared understanding regarding human rights has emerged in Europe (or world-wide). As an example, we can take the case of same-sex marriage. While in the past there may have been no agreement across countries around the question of whether same-sex couples should have the right to marry, when deciding a case on whether such a right exists, ECtHR judges can examine rules, laws and practices of member states of the CoE today or other organizations, such as the European Union, and determine that it is the case that new norms around the acceptance of same-sex marriage have emerged, justifying a decision to create such a right within the context of the ECtHR. Thus far, EuC analysis has led the ECtHR to conclude that no right to marriage shall exist for same-sex couples in Europe (e.g., Hämäläinen v. Finland [GC], 16-7-2014, 37359/09), but also that such couples shall be afforded some form of legal recognition of their union, such as civil law partnership (Oliari and Others v. Italy, 21-7-2015, 18766/11 and 36030/11).

Within the EuC framework, the ECtHR may also engage with comparative analysis of sources of law outside of the practice of the CoE member states. For instance, it can consider the practice of international organizations, such as the CoE itself or the European Union (EU), and consider human rights standards developed by other, non-European institutions (e.g., the United Nations (UN)) or even by non-European states. If the ECtHR identifies the existence of consensus, however defined, on a specific human rights issue, it will then move on to recognize

pan-European standards that are binding on all states under its jurisdiction. In the absence of consensus, states and their national authorities enjoy a wider margin of appreciation[10], that is discretion to make their own regulatory choices. To that end, the ECtHR's level of judicial scrutiny is lower.

Although we can offer a basic definition of EuC, as we have here, much about the method and its use by the Court remains unknown. A leading academic volume on the topic has stated that:

> [T]he designer and game master of the EuC 'game' [the ECtHR] has not (yet) provided a full, clear, and detailed 'manual' on the definition of consensus and on the rules and outputs of the game. [...] [A]lthough it is evident that a considerable number of states must have developed a common practice for EuC to emerge, it is unclear where the cut-off point is (i.e., how many states will be required to reach the 'consensus' threshold). In the same vein, no detailed explanation has been provided regarding the different 'shades' of consensus and how these may correspond to the different terms that make up the 'consensus lexicon' employed by the ECtHR (Kapotas and Tzevelekos 2019, 9–10).

The authors speculate that the ECtHR judges may, themselves, lack a common understanding as to what EuC is, and perhaps also a common vocabulary. The Court will often use different terminology to refer to the same or similar concepts, making it difficult to ascertain whether the Court is truly referring to EuC. Alternatively, the authors suggest that our ignorance about the nature of EuC may be the result of a strategic choice on the part of the Court, for whom the lack

---

[10] Specifically, the term "margin of appreciation" refers to the "room for manoeuvre the Strasbourg organs are prepared to accord national authorities in fulfilling their obligations under the European Convention on Human Rights" (Greer 2000).

of a detailed EuC rule book may increase its flexibility when employing the method (see Kapotas and Tzevelekos 2019, 10). While we can be confident that the ECtHR uses EuC in certain judgments, and legal scholars have identified some language associated with the use of EuC, there remains a significant lack of clarity about the nature of this important legal method. This lack of clarity stands in contrast to research on other, related legal methods of interpretation, such as "margin of appreciation," which have clearer definitions and are more obvious when employed by the Court.

In our analysis, we treat the EuC as a latent variable that we must measure, much like other latent concepts that social scientists often measure with measurement models (e.g., democracy, rule of law, and ideology). We aim to empirically uncover the nature of EuC using the texts of the ECtHR's judgments. We hope to provide the "manual" that the court has not yet offered up by using the information in the texts of judgments to uncover both explicit and latent use of EuC in the Court's decision making.

The Law and Politics of ECtHR Decision-Making

Political scientists often view decision-making processes to be just as important as the outcomes of those processes – e.g., when assessing the state of democracy (Coppedge 2002; Munck and Verkuilen 2002). Process and nature of argumentation matter because they set precedent for how future decisions are made, which ultimately impacts policy outcomes (see e.g., Riker 1986), and also because they may affect views regarding the legitimacy of a decision. In the context of legislative decision-making, it is common to focus on rules and procedure, both formal and informal, to understand decision-making (see e.g. Cox and McCubbins 2005, 2007; Reh et al. 2013; Shepsle 1979). When studying courts, political scientists are less likely to focus

on how courts make decisions and the nature of legal argumentation that they use. Instead, their focus has been on the outcome of the judicial decision-making process – e.g., the votes of judges, themselves (see e.g. Epstein, Landes, and Posner 2012; Martin and Quinn 2002; Segal and Spaeth 1993; Voeten 2008), and the policy impact of the judicial rulings (Hafner-Burton and Tsutsui 2005; Mathew D. McCubbins, Noll, and Weingast 1987; Matthew D. McCubbins, Noll, and Weingast 1989).

Legal scholars, in contrast, are often just as interested in the legal "tools" and procedures that judges and courts use to reach their decisions and to develop their arguments as they are in the decisions themselves (Kapotas and Tzevelekos 2019). The nature of legal argumentation can be both of intrinsic interest and potentially influence the level of legitimacy of the ruling within the legal community. Unlike in the contexts where political scientists have frequently studied rules and procedures, such as legislatures where the rules of procedures and decision-making tools are often written down, in the judicial context, the nature of the legal methods of interpretation used by judges to make decisions are not always clear and their use is more likely to be informal. Courts do not always provide detailed reasoning, and when they do, they do not always explicitly label or name the interpretative method that they have employed. This leaves room for legal scholars to debate the nature of the legal decision-making process and the nature of the legal methods employed by judges. However, it is difficult to study these legal decision-making tools systemically (but see Howard and Segal 2002, Helfer and Voeten 2021; Stone Sweet, Sandholtz, and Andenas 2021).

The empirical literature on the ECtHR, and international courts more generally, comes mostly (but not entirely) out of the political science tradition and is outcome-oriented. Generally speaking, it fits into one of the following categories: the study of judicial decisions and the

prediction of outcomes (Aletras et al. 2016; Medvedeva, Vols, and Wieling 2020; Voeten 2008, 2021), the explanation of variation in compliance of judgements (Grewal and Voeten 2015; Hillebrecht 2009, 2012, 2014; Panke 2020; Stiansen 2019, 2021), and the study of citizens' support for courts and the effect that support has on court rulings and compliance (Cichowski 2006; Dinas and Gonzalez-Ocantos 2021; Madsen 2020; Stiansen and Voeten 2018).[11]

A number of studies focus on the political motivations of judges and the possible biases that may shape their decision-making (Voeten 2007, 2008, 2021). The literature looks at correlations between the characteristics of appointed judges (e.g., gender, country of origin, legal system of origin, and previous career history) and the outcomes of cases. The results of these studies are quite mixed. Female judges are more likely to find in favor of cases filed by women, but only on discrimination cases (all else equal), which fits into the theories that suggest that female judges have unique experiences on discrimination relative to men, making them potentially more sympathetic on issues related to discrimination (Moyer and Haire 2015; Palmer 2001; Voeten 2021). Judges sometimes are biased towards their country of origin, though a judge does not want to be a "lone dissenter," and it depends on their country of origin (Voeten 2008). There is no evidence that career history or legal origin play a role in judicial outcomes (Voeten 2008). There seems to be few sources of bias for judges of the ECtHR, leading Voeten (2008) to claim that, "the overall picture is mostly positive for the possibility of impartial review of government behavior by judges on an international court...ECtHR judges are politically motivated actors in the sense they have policy preferences on how to best apply abstract human rights in concrete cases" (Voeten 2008, 417).

---

[11] But see Lupu and Voeten (2012) and Helfer and Voeten (2014) for exceptions.

A second recent set of work within comparative judicial politics has started to use machine learning methods to study and predict the outcomes of ECtHR cases. Aletras et al. (2016) were one of the first to use the text of the ECtHR judgments as data to build a machine learning classifier to predict the outcomes of the cases. Their model, as well as some of their text preparation procedures, were improved upon by Medvedeva, Vols, and Wieling (2020). This research shows that, with some degree of certainty, machine learning tools can be applied to the texts of ECtHR judgments to predict judicial outcomes.[12]

A third strand of literature explains variation in compliance of judicial decisions. As the ECtHR is an international court, its enforcement powers are significantly lower than that of a national court. As a result, non-compliance has varied over time and by country (Panke 2020). The ECtHR has to rely on domestic institutions for compliance with its judgment (which is monitored by a political body within the CoE, namely the Committee of Ministers), so compliance is found to be higher in countries with strong domestic institutions combined with countries that have a desire to protect human rights (Hillebrecht 2012, 2014). Alternatively, new democracies have an incentive to signal to their counterparts that they prioritize human rights, so compliance is slightly higher in new democracies (Grewal and Voeten 2015). Compliance is also found to be higher in cases where the ECtHR provides remedial indications, which are specific requirements for the compliance of a case (Stiansen 2021). Regardless of type of country, if there are legislative changes necessary for compliance, which makes up about 25% of cases, then compliance will be lower (Stiansen 2019).

---

[12] Neither Aletras et al. (2016) nor Medvedeva, Vols, and Wieling (2020) treat the cases as one cohesive document-both only look at specific parts of the case to use with the classifier in order to reduce noise. Medvedeva, Vols, and Wieling (2020) have slightly higher precision than Aletras et al. (2016), which they argue is because they remove the Law Section (and find The Facts section is the most important for the classifier). Overall, Medvedeva, Vols, and Wieling (2020) find around 75% accuracy when taking a random sample to train and test the data.

Lastly, literature on the ECtHR tries to explain variation in levels of public support for the Court. Public backlash seems to be at its highest when the agendas of the international institutions are in conflict with the agendas of the domestic institutions (Madsen 2020), and this could have an impact on judicial appointments to the court. Stiansen and Voeten (2018) find that in countries where backlash is higher, governments tend to appoint more restrained judges. These studies use public opinion data and experimental survey designs to understand the conditions of support for the Court. Here, the work distinguishes between non-compliance and backlash, with backlash viewed as the more serious threat as it questions the legitimacy of the court (Dinas and Gonzalez-Ocantos 2021). Overall, this research finds that invoking a democratization argument, namely that the Court helps to improve democratic standards, does not increase public support for the Court. Instead, public support only increases when the court is framed as an agent of a member state (Dinas and Gonzalez-Ocantos 2021).

None of this literature has explored the nature of legal argumentation in ECtHR decision-making or its relationship to any of the outcomes of interest to these studies. While we do not explicitly look at the relationship between EuC and outcomes with respect to compliance or legitimacy in our work either, we argue that we take a first step – the conceptualization of EuC – towards understanding how legal reasoning and the use of this particular method of legal interpretation can impact outcomes of interest to both quantitative social scientists and legal scholars interested in the ECtHR.

Why Study EuC?

One might reasonably ask why we bother to study the tools of legal decision-making used by ECtHR judges at all. After all, citizens are unlikely to understand or care about the tools of

legal decision-making and therefore the use of one tool over another is highly unlikely to affect public opinion or public backlash towards the court. Nor are governments or elected politicians likely to take the logic of judicial decision-making into account when making political decisions about whether to comply with a court decision. We argue that there are nevertheless very good reasons to study EuC and to develop an understanding of how judges use the tools of legal decision-making at their disposal.

First, we intrinsically care about and are interested in how elites make decisions and how they justify their own decisions to themselves (even if these justifications do not directly impact outcomes). Elites, including ECtHR Judges, are fundamentally different from the general population in the way that they process information (Hafner-Burton 2021). This means that studies about how general populations make decisions cannot be used to study the decision-making of elites (Hafner-Burton 2021). At the same time, given their position, it is difficult to survey elites about their decision-making in the same way that we can survey the public. Even when elites can be tracked down to answer questions, they may be cautious about sharing information about how they reach their decisions (Hafner-Burton 2021; Hafner-Burton, Hughes, and Victor 2013). We may be able to gain insight into the reasoning of judges by examining the texts of their judgments.

Second, even if the ECtHR's use of specific decision-making tools, and EuC in particular, are unlikely to have a direct effect on public opinion[13] or change the behavior of governments, they may influence the views and decisions of legal experts both at the international and national levels. At the international level, human rights lawyers, who bring and argue cases before the court, may care about how judges have arrived at a legal decision, and it

---

[13] Experimental research has shown however, that citizens are more likely to reject policies if told that these policies would violate international law (Tomz 2008).

may shape whether they bring cases and how they argue them when they do. Indeed, existing research has demonstrated that the reasoning of international courts has "persuasive authority", beyond the behavior of compliance constituencies (Helfer and Voeten 2014). By fostering alliances between international and national judges it may influence national judges to use ECtHR judgements in their reasoning (Koshla 2011). The use of EuC – specifically – can confer legitimacy to a judgment, with the Court effectively saying in its reasoning that the states, themselves, have set the standard, and not the court. When a "consensus" exists, a state found to be in violation of the ECtHR is said to be in violation of norms that are upheld by substantial number of neighboring European states. In other words, the Court is saying that it is not telling a state what to do based on standards that the court has decided to set ex cathedra, but rather it is the states' own neighbors who have developed and adopted new norms. This is similar to the mechanism of "emulation" in research on transnational policy diffusion. Emulation can be described as a process of diffusion of norms between nation states that is based not on their objective properties but rather on their "normative and socially constructed characteristics" (Gilardi 2012, 461). If norms have become broadly accepted or even internalized, national actors are more likely to support them because noncompliance might have negative consequences for their national legitimation (Finnemore and Sikkink 1998; Gilardi 2012).

Finally, compliance with, and the impact of, the ECtHR is not merely a function of actions taken by elected politicians and governments when making policy decisions. The ECtHR must be incorporated into the national legal systems through national law. In the UK, for example, incorporation happened through the Human Rights Act of 1998. As such, national judges upholding national law can look to text of ECtHR judgments for an understanding of how to apply the ECtHR, which is also a part of their own national law. The ECtHR's quality of legal

reasoning and the legal methods of interpretation used may impact legitimacy in the eyes of national court judges and lawyers and may impact how they apply it in their own legal reasoning when making judgments at home. These domestic judgements may not only influence national political actors but, in turn, also judicial reasoning at the ECtHR, thereby creating a feedback loop in the creation of norms from international courts to national courts and political processes – and back to the international level.

Hypotheses

Although our primary research is descriptive, we lay out some basic hypotheses that we seek to examine with respect to the changing use of EuC by the Court. First, we hypothesize that the use of EuC has increased over time. Much of the qualitative legal literature on EuC has asserted so much (e.g., Dzehtsiarou 2015, Kapotas and Tzevelekos 2019) and we seek to test this simple proposition more formally here. EuC is, after all, a legal tool that the Court has invented for itself. As the tool has become more accepted by Judges its use should increase.

H1: The use of EuC increases over time.

Second, we hypothesize that EuC should be used more in cases where certain states are respondent, and where certain articles of the ECHR are at issue. EuC is specifically used when cases involve unsettled, sensitive, and difficult moral issues, such as same-sex marriage or abortion. We would not expect it to be used in cases where the Court has already established key standards (e.g., the absolute nature of the prohibition of torture). Thus, the cases where these we expect to find more EuC are those involving Western European democracies (which are less likely to engage in state torture, but more likely to face questions surrounding difficult moral

issues) as respondent states and those involving articles of the ECHR that touch on moral and sensitive issues, e.g., Art 8 "Right to Respect for Private and Family Life."

H2a: EuC is more likely to be used in cases involving western European democracies as respondent states.

H2b: EuC is more likely to be used in cases involving treaty articles that cover sensitive moral and legal issues.

Research Design

To study the nature of judicial decision-making by the ECtHR, we analyze the legal texts that the Court produces, namely the judgments of the Grand Chamber. Our analysis is limited to Grand Chamber judgments that run through the end of 2019, at which time there were 465 finalized Grand Chamber cases. By focusing on Grand Chamber judgments, we cover the most important decisions made by the ECtHR, although we plan to eventually expand our analysis to cover Chamber judgments, as well. To measure and quantify the Court's use of the EuC method, we use tools from both law and computational social science. Our analysis consists of six steps.

Our process is roughly as follows:

1. *Identify the parts of the ECtHR judgments where the ECtHR lays out its comparative data and legal reasoning and which could potentially contain EuC reasoning.*

2. *Conceptualize EuC and develop a Human Coding scheme on the basis of our current understanding of EuC*

3. *Hand code a random sample of Grand Chamber cases*

4. *Train a classifier to uncover other possible instances of EuC*

5. *Examine correlates of the Court's use of EuC language: e.g., time, the government involved, the Articles of the ECHR in question (i.e., the human rights at issue), the judges hearing the case, etc.*

We explain each of these steps in turn.

Identify the relevant parts of the ECtHR judgments: Similar to previous work looking at ECtHR case documents, we do not analyze the entire case. Specifically, we restrict the case documents to The Law Sections and The Relevant Law and Materials sub-section (which sometimes appears in The Procedure or The Facts sections) as these are the parts of the cases that contain the legal argumentation as laid out by the Court. Additionally, we are only interested in analyzing parts of the cases where the Court is "speaking". Within a case, there are arguments from both sides as well as the analysis from the Court. EuC reasoning can only logically be present in the section of the judgment where the Court is developing its own reasoning rather that reciting the arguments of the parties to the dispute or of interventions by third-parties. To identify when the Court is speaking, we analyzed the structure of the cases as well as built a machine learning classifier based on the language the Court used in the Grand Chamber.[14]

Conceptualize EuC and develop a plan for Human Coding: We first require significant input from humans, namely highly-trained human rights lawyers, to conceptualize what is EuC language. We rely upon our team of academic lawyers who specialize in European human rights law to identify certain language and patterns indicative of the EuC method. Nevertheless, a significant challenge facing our research is that even these highly trained human rights lawyers who study decision-making in the ECtHR do not always agree on what constitutes the use of EuC by the Court. This problem is not unique to the study of EuC but crops up whenever

---

[14] This process is described in more detail in the Appendix.

studying and attempting to measure imprecisely defined latent concepts (e.g., democracy, hate speech, populist rhetoric to name a few). While some instances of EuC language are easy to identify, even for non-trained coders, there are instances where even highly trained lawyers do not agree.

We tackle this problem in two ways. First, we decide to focus on the occurrences of *EuC language* rather than the use of EuC, itself. The difference is subtle; while in most instances, when the Court uses language associated with EuC, it is, in fact, using EuC reasoning. However, there may be instances in which language is used but not necessarily to the tool of EuC, itself. The lawyers in the project are more readily able to agree on what constitutes EuC language than whether the Court uses EuC, itself, in specific instances. Second, we develop a tagging scheme to highlight this language at the paragraph level within the judgments.

To identify EuC language, we focus on identifying two factors: one, references to external sources within the cases; and two, the use of EuC language. First, it is important to identify references to external sources as sources that are extraneous or external to ECtHR text a necessary, but not sufficient, precondition for the use of EuC. The EuC method draws on domestic human rights standards (set by domestic courts or by national law makers for instance), the practice of international institutions, or even the practice of states other than the ECHR parties for various comparative analysis purposes, including to explore whether these sources and instances of practice of states or organizations (such as the EU) justify setting a common European standard that should be the practice of all ECHR parties as well. To do this, not only did we identify external sources, but we identified types of external sources. These sources ranged from domestic practice of the ECHR parties, domestic practice of third states, sources stemming from EU legal order, sources stemming from the CoE system outside of ECtHR text

and case law, other sources including hard or soft international law or case law of third international courts, and others (such as scientific evidence or a bibliography). Together, these forms of external sources make up the universe of "ingredients" for consensus to be identified – yet they are a necessary but not a sufficient condition, because references to such sources can be done for different (albeit possibly contiguous) interpretive purposes besides EuC.

After identifying references to external sources, which are a necessary precondition for EuC, the lawyers focused on identifying EuC language by tagging paragraphs within judgments as follows: Tag 1 is used for the explicit use of the term "consensus" by the Court itself, which would (almost) always indicate that the judges are, indeed, employing consensus. We tag the word alone and when used on its own or in conjunction with other terms (e.g., "European consensus", "Scientific consensus", "emerging consensus", and "international consensus"). The second category -- Tag 2 -- indicates the use of quasi-explicit language and is a little more nuanced than Tag 1, but still clearly indicative to lawyers that the Court is employing EuC language. Examples of Tag 2 language include phrases such as "the vast majority of [states/countries/member states/contracting parties]", "no uniform approach", "trend", "a significant number of [states/countries/member states/contracting parties]". These phrases indicate that the Court is engaging in comparative analysis of the practice of the ECHR state parties or of other states or more generally types of sources (e.g., international law standards) to determine whether a new consensus around the existence of a human right is emerging. The final tag -- Tag 3 -- is given to paragraphs when Tag 1 or Tag 2 language is absent, but the Court refers to extraneous sources and the context suggests to the lawyer-coders that the Court could be using EuC logic. Tag 3 paragraphs were then examined by the whole team of lawyers and further broken down into those paragraphs where the lawyers were more confident that EuC is being

employed and those where it is not. Only those paragraphs where higher certainty exists retained their Tag 3 coding. This tagging scheme was developed by the team of lawyers after a careful reading of a random sample of Grand Chamber cases.

Hand code a random sample of Grand Chamber cases: Having developed the coding scheme, the team of lawyer-coders applied it to a random sample of 237 of the Grand Chamber judgments. To ensure the highest ease of functionality across our team, we did all the hand coding in Microsoft Word with the use of highlighting and commenting functions. Then, we imported these documents in *R* for text processing and quantitative analyses. We found that this process worked well across different skill levels, and it allowed the structure of the documents to be preserved (such as the different section titles, font size, indentation, etc.).

Train and run a classifier to uncover possible instances of EuC: We build a classifier to examine whether we can uncover, based upon systematic use of language, instances of paragraphs that lawyers would agree could be tagged as containing consensus language. We start by training the classifier on the sample of hand-coded texts.[15] We then refine it and run it out-of-sample on the whole Grand Chamber corpus. We sample the paragraphs that the classifier has indicated to contain consensus. These paragraphs were then given to the lawyers to determine whether they should be coded as containing any type of consensus language. Each paragraph was given to multiple lawyers to code. Coding was blind so that lawyers did not know how other lawyers had coded the same paragraph. This information is then used to further refine the classifier, with the goal of applying it out-of-sample to the whole Grand Chamber.

Examine correlates of the Court's use of EuC language. In a final step, we can use the results of the analyses to identify correlates of EuC, including those in the hypotheses stated above.

---

[15] The Appendix contain information on how we preprocessed text, the classification algorithms, and more details on the machine learning classifier.

Scholars of the ECtHR have noted that the Court's use of EuC has increased over time. Others have hypothesized that the Court may be more likely to use the tool when cases involve certain member states. For example, the Court may use EuC to demonstrate to certain more skeptical or recalcitrant governments (e.g., the United Kingdom) that many or most other states hold a different view about the human right in question. Likewise, certain judges or judges from certain states may be more likely to engage in EuC reasoning. And EuC might be used more with regard certain human rights than others. In this research, we focus on the results from the hand-coded sample as the classifier results still yield too many false positives – tagging language as consensus when it is not.

Findings

As described above, our first task was to develop our coding scheme and apply it to a sample of Grand Chamber judgments. We use the tagged text from this human-coding process to train our classifier. While the classifier analysis provides some insights into our research questions and hypotheses, it identifies to much text as containing consensus when, in fact, it does not. We uncover a large number of false positives in our classifier results. Thus, we use these human-coded cases to investigate our hypotheses.

Classifier Analysis

We start training our machine learning classifier using the hand-coded cases. While these cases do not represent the entirety of all the Grand Chamber cases, they are a random sample and sufficiently representative to understand general trends and overall performance of the classifier. As described in more detail in the Appendix, we run multiple different classifiers with various forms of text preprocessing and transformation. We run all analysis at the level of the paragraph.

Due to the strong imbalance of consensus labelled paragraphs to non-consensus labelled paragraphs, we up-sample consensus language and down-sample non-consensus language in our training set. Ultimately, this leads us to a training set of a ratio of around 1:2 (consensus to non-consensus language paragraphs), which we randomize into 10 different training and test sets to run the different machine learning classifiers. This allows us to get a baseline performance of the different models while ensuring that specific paragraphs are not driving the results.

First, we discuss the overall performance of the different machine learning classifiers. The classifier includes all tags in the training and test sets (Tag 1, Tag 2, and Tag 3) and it tries to predict whether a paragraph contains consensus or not, with EuC language operationalized as a binary variable. The best performing model is a support-vector machine model (SVM) using unigrams and the least amount of text pre-processing. For the best performing model, our F-1 score, an overall performance measure which ranges from 0 to 1, is 0.58, which is rather low. Despite this low F-1 score, our overall specificity scores are high (0.98). This is encouraging as it means that if the classifier identifies a paragraph as consensus, it is nearly always contains consensus language as identified by the lawyers. This high specificity score is coupled with a low precision score, meaning that there are also a high number of false positives -- non-consensus language identified by the classifier as consensus. It is these false positives that drive down our F-1 scores. With the aim to improve the classifier, specifically to increase the F-1 and precision metrics, we attempted to add more information to the machine learning classifier in the form of other bigrams or word pairs, but this did not result in better classification.[16]

---

[16] Specifically, we ran Wordscores from Laver, Benoit, and Garry (2003) to identify bigrams that were associated with our consensus tags. We ran the Wordscores using bigrams for the Tag 1s; Tag 2s; Tag 3s; and Tag 1s, Tag 2s, and Tag 3s. We ran the Wordscores on a 0 to 1 scale, taking a random sample of non-consensus language for our reference text assigned a 0, and we used a 1:2 ratio of consensus to non-consensus language. From each sample of consensus language, we took the bigrams that scored a 0.7 or higher. This in total lead to around 85 bigrams from

We can also look at the specificity of the classifier aggregated to the case level. In the vast majority of cases, the classifier accurately uncovers 100% of the consensus paragraphs at the case level. There is one case where the classifier predicts 0% of the consensus paragraphs, but this case has one instance of tagged consensus language. This confirms that the classifier is very good at accurately tagging instances of consensus language. Nevertheless, the problem of false positives becomes even more apparent at the case level since virtually every case has at least one paragraph that is identified as containing by the classifier. If we use the aggregation rule that if at least one paragraph contains consensus the judgment uses consensus reasoning – the rule that we use in the hand-coded sample – we would find that virtually every case uses the consensus reasoning – something that is simply not the case.

Ultimately, the reason behind the low level of precision is the relatively small amount of data when it comes to consensus language. There are relatively few paragraphs that contain EuC language, which both means that there is not much data that the classifier can use to identify consensus, and also that it is very easy to err on the side of overclassifying consensus, simply because there are so many paragraphs that do not contain consensus compared with those that do. Asking the algorithm to find consensus turns out to be akin to asking it to find the proverbial needle in a haystack.

---

the Tag 1s, Tag 2s, and Tag 3s. Examples of Tag 1 bigrams are: contract state, member state, margin of (total of 14 bigrams). Examples of Tag 2 bigrams are: state in, a number, contract state, practice of (total of 49 bigrams). Examples of Tag 3 bigrams are: in country, in state, member state, country a (total of 18 bigrams). When running the Tag 1s, Tag 2s, and Tag 3s together, there are a total of 79 bigrams. We added these bigrams back into the document frequency matrix and ran the SVM machine learning classifier again with the Wordscores bigrams as features with the aim of adding more information to improve the overall performance of the classifier. In general, while some combination of features may improve some metrics, the results do not change much from the classifier run without features. Overall, adding the bigrams does not improve the performance metrics of the classifier.

Hand-Coded Sample

The low level of precision and correspondingly high number of false positives from our classifier analysis makes it difficult to use the classifier results to evaluate our hypotheses. The classifier results are not yet sufficiently accurate for us to determine whether a particular paragraph contains EuC language. Thus, to examine whether the Court uses more EuC language over time, or when handling cases with certain states as respondents, or with respect to certain articles, we return to our hand-coded sample. Here we can say with a very high degree of confidence that we have identified all consensus language and associated consensus reasoning, without tagging cases or paragraphs that do not contain consensus language.

We begin by showing the increase in EuC over time at the case level. Figure 1 shows the increase in cases containing EuC language as a percentage of all cases in the Grand Chamber (among those in our hand-coded sample) by year. There has been a clear positive trend over time. Prior to the early to mid 1990s, judges used almost no EuC language in their judgments, however in the last decade EuC language appears in up approximately 40% of Grand Chamber cases per year. We see a similar time trend if we look at the percentage of consensus cases by individual respondent states (e.g., the UK).
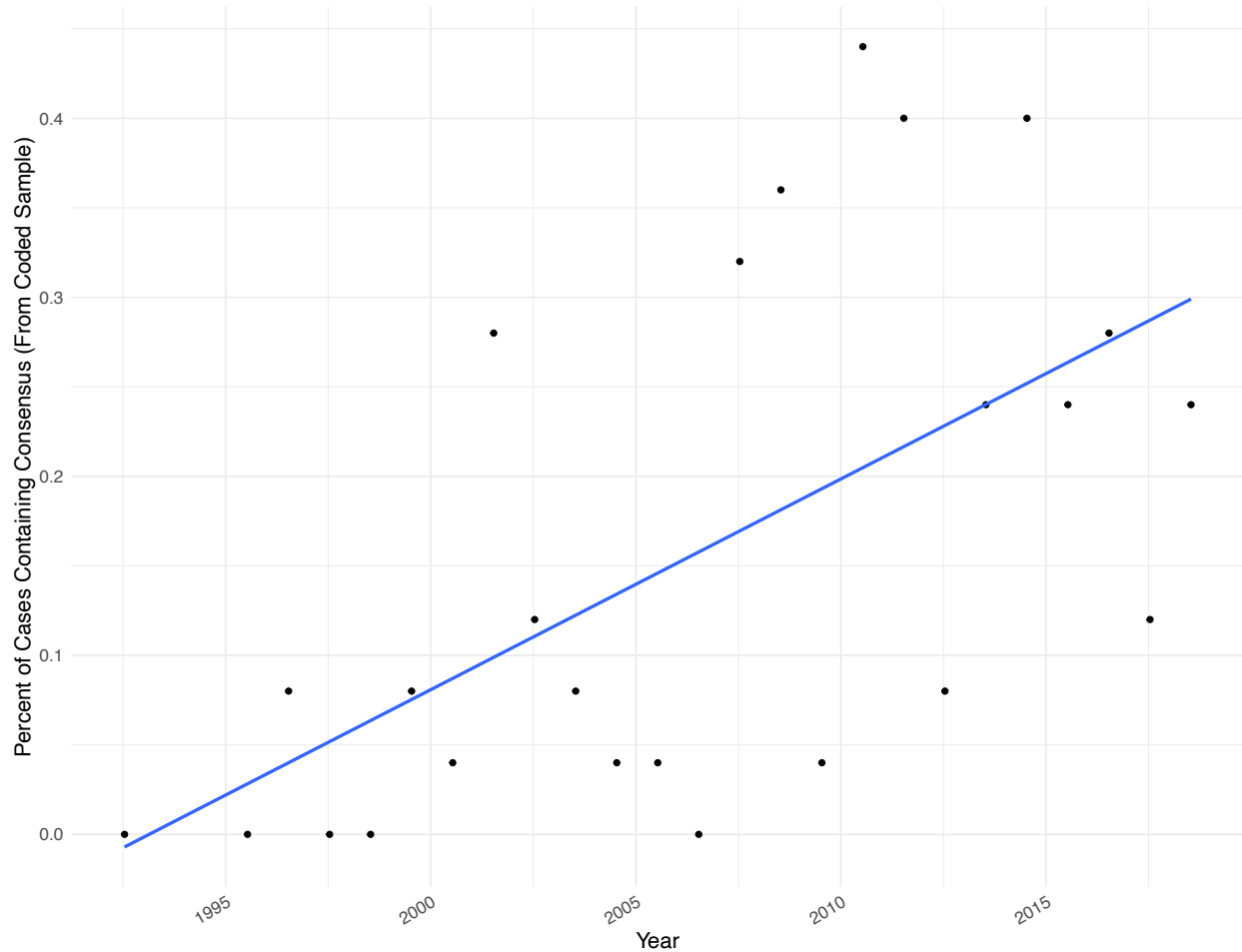
Figure 1: Increase in the Use of Consensus Language Over Time (Hand-Coded Sample), Percent of Cases per Year Containing Consensus Language.

Next, we show how the Court uses EuC when deciding cases with certain states as respondent states. Again, using the hand-coded data, Figure 2 shows the number of total coded cases and the number of cases containing consensus by respondent state. Here we clearly see that some states are more frequently on the receiving end of EuC reasoning than others. The UK is the state against which EuC reasoning is most often used. It is also involved in a large number of cases, but even as percentage, the number of judgments using EuC language remains high. Other western European states follow behind the UK. Most notably, these include Italy and France. Surprisingly, perhaps, Turkey is also subject to a high number of consensus judgments. On the

whole, the hypothesis that western European states are often on the receiving end of consensus judgments finds support.
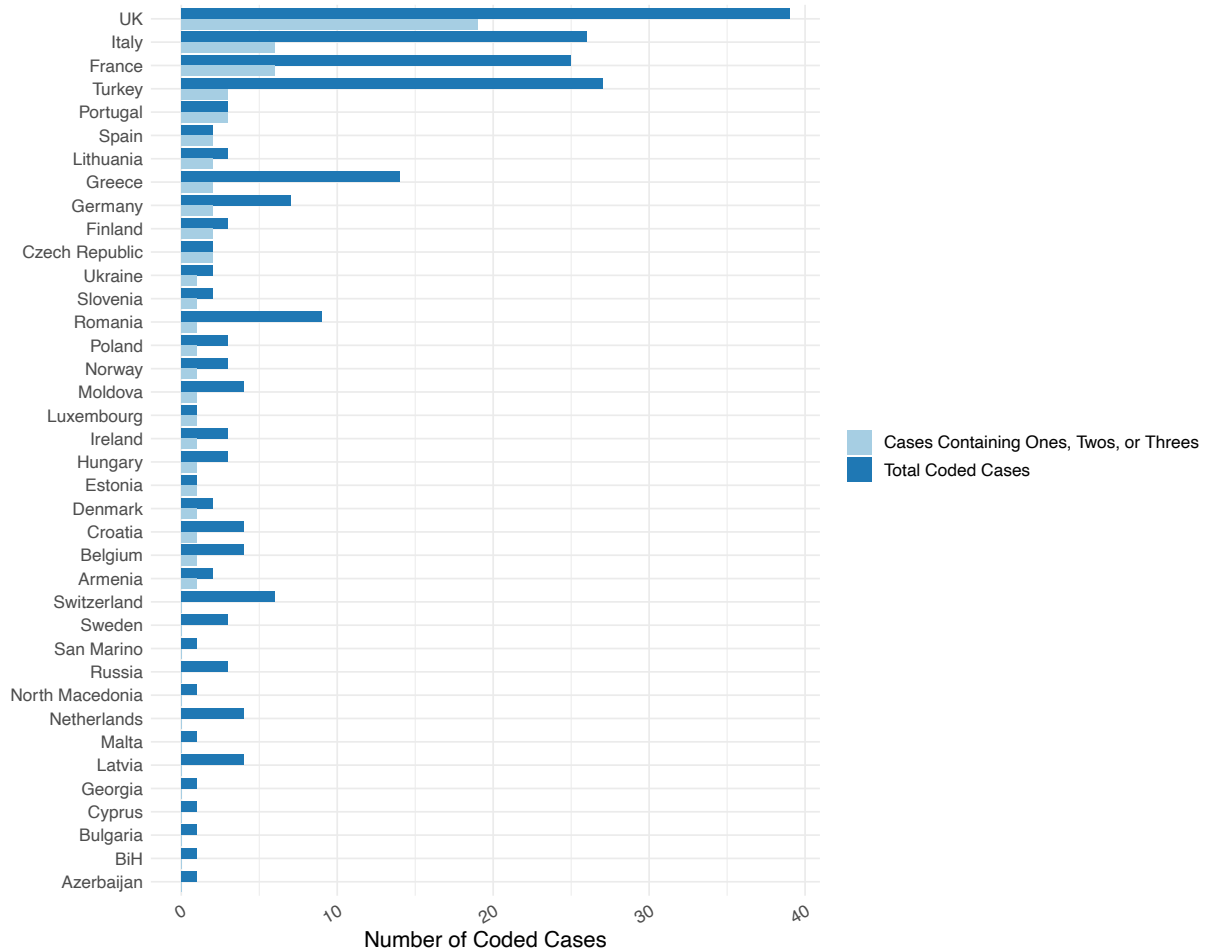

Figure 2: Number of Total Cases in Which a State is the Respondent (Dark Blue) and the Number of Cases Containing Consensus in Which a State is the Respondent (Light Blue). Hand-Coded Sample Only.

Next, we examine which ECHR articles are most frequently associated with EuC language within our hand-coded sample. The results are presented in Figure 3. Here we see that in our hand-coded sample, Article 8 is the ECHR article most associated with consensus language, followed by articles 6, 10, and 14. These articles cover "Right to Respect for Private and Family Life" (Art. 8), "Right to a Fair Trial" (Art. 6), "Freedom of Expression" (Art. 10) and

"Prohibition of Discrimination" (Art. 14). Clearly these articles cover areas of law that touch on sensitive moral and societal issues, and around which new consensus could potential emerge.
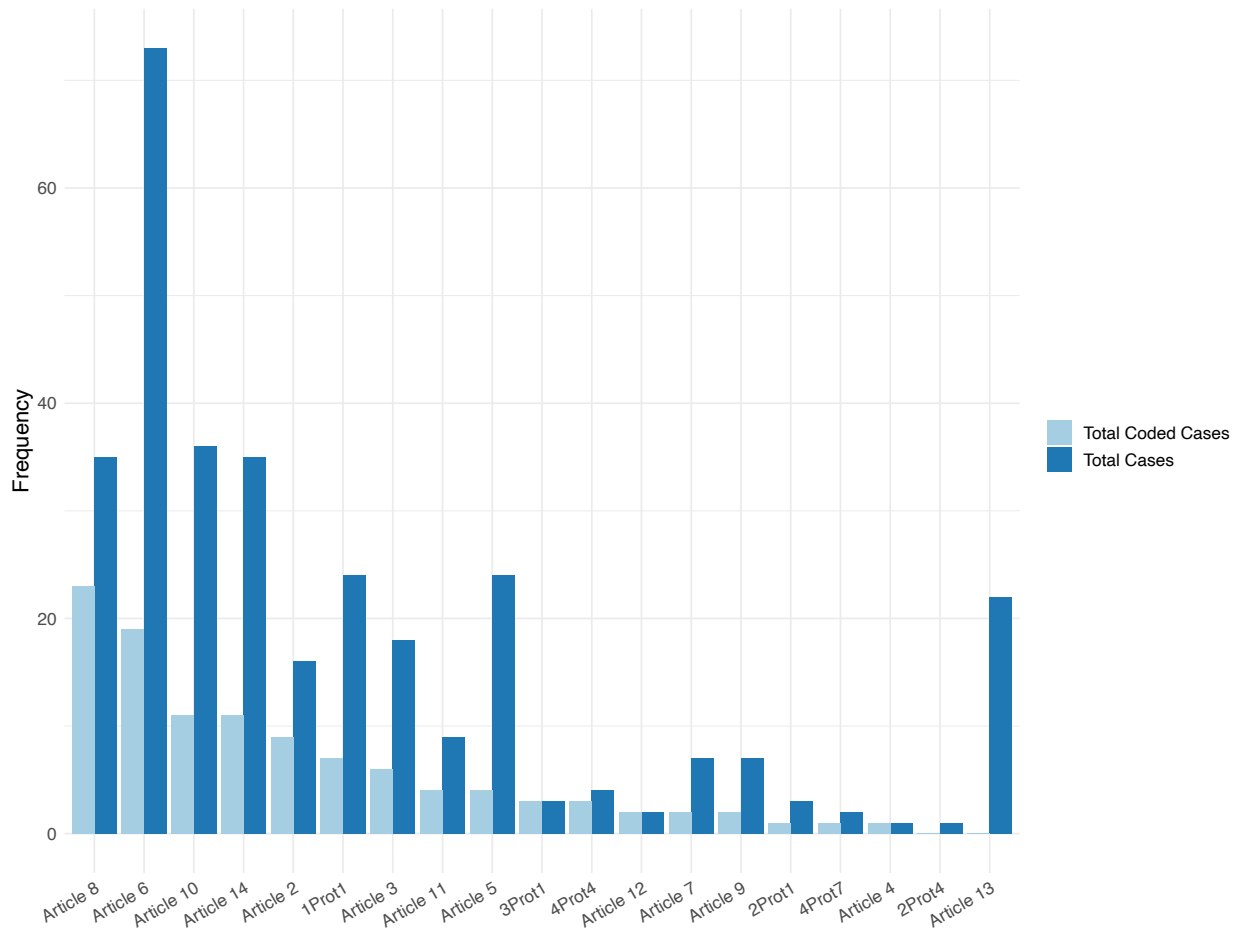

Figure 3: Number of Cases in Involving Particular Articles and Number of Cases in Which Consensus Language is Used with Respect to Those Articles (Hand-Coded Sample).

Finally, we can use our hand-coded sample to examine whose states' practice is considered by the Court when it explicitly names the states whose practice inform its consensus analysis. We can view these states as those whose practices the Court is most likely to take into account when building an argument for the purposes of consensus analysis. Figure 4 shows the number of hand-coded cases containing consensus language in which a particular state is mentioned in the paragraphs containing consensus language. We determine these counts by

combining our hand-coded tagging results identifying consensus with our hand-coded highlighting of the different sources used by the Court for comparative analysis. We take all highlighted text containing sources referring to countries' practices and pass a dictionary of country names over it. By doing so, we identify which countries are mentioned by the Court with respect to consensus.
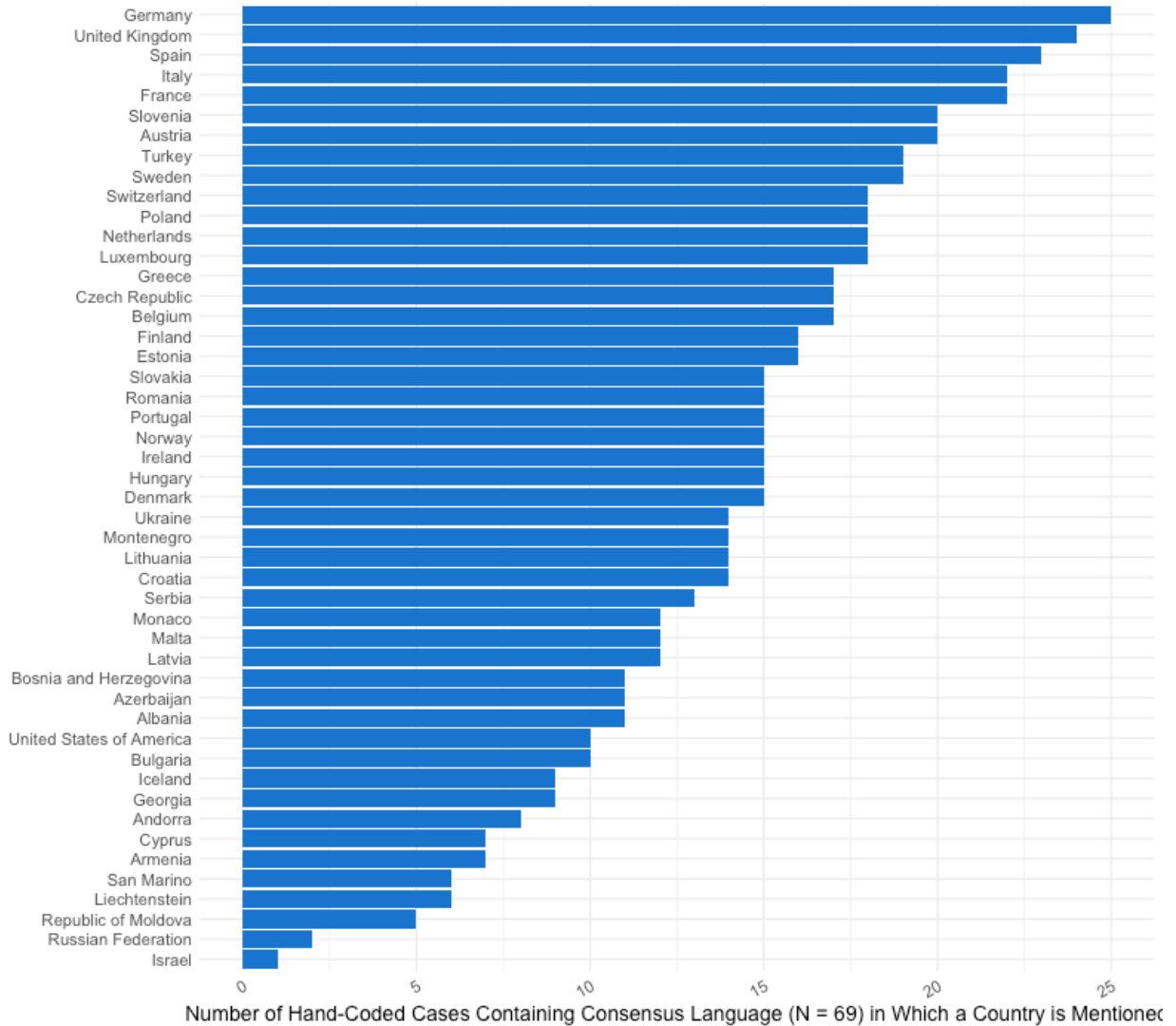

Figure 4: Number of Consensus Cases in Which a Country is Mentioned with Respect to Consensus Language (Hand-Coded Sample)

We find that Germany, the UK, Spain, Italy and France top the list. This is not surprising as they are all large Western European states. However, after these top five many different states

are mentioned, including Slovenia, Turkey, Poland and others. In general, it appears that Court draws on a wide variety of state practice when citing countries' laws to determine if consensus exists. While it clearly refers quite often to the practices of large, Western European states, it by no means focuses exclusively on Northern Europe. It is worth noting that the only two non-signatories of the ECHR to receive mentions are the United States and Israel. The former is being considered more by the ECtHR in the framework of consensus analysis than the Russian Federation, which means that the US is more impactful in shaping European Human Rights Standards than Russia.

Lastly, we can look at how the Court builds consensus arguments with respect to one particular member state – the UK. The UK is interesting to examine in greater depth for several reasons. First, it is both the most frequent respondent state in general before the GC, and also the state most subject to consensus reasoning by the Court when it is a respondent. Second, it has a unique relationship to the Court in that it both pressed hard to establish the Court and to build a human rights regime across Europe in the initial decades following the Second World War, but it also has criticised the Court in more recent years, arguing that the ECHR and the Court, itself, encroaches on its sovereignty. This has resulted in the UK refusing to comply with certain ECtHR judgments against it (e.g. voting rights of prisoners). Figure 5 is the same as Figure 4 but focuses solely on those cases where the UK is a respondent state. Again, we see both that large, Western states, such as Spain, Germany, and France are most frequently mentioned by the Court, but many other states are cited as well. The Court cites the practices of a wide array of countries
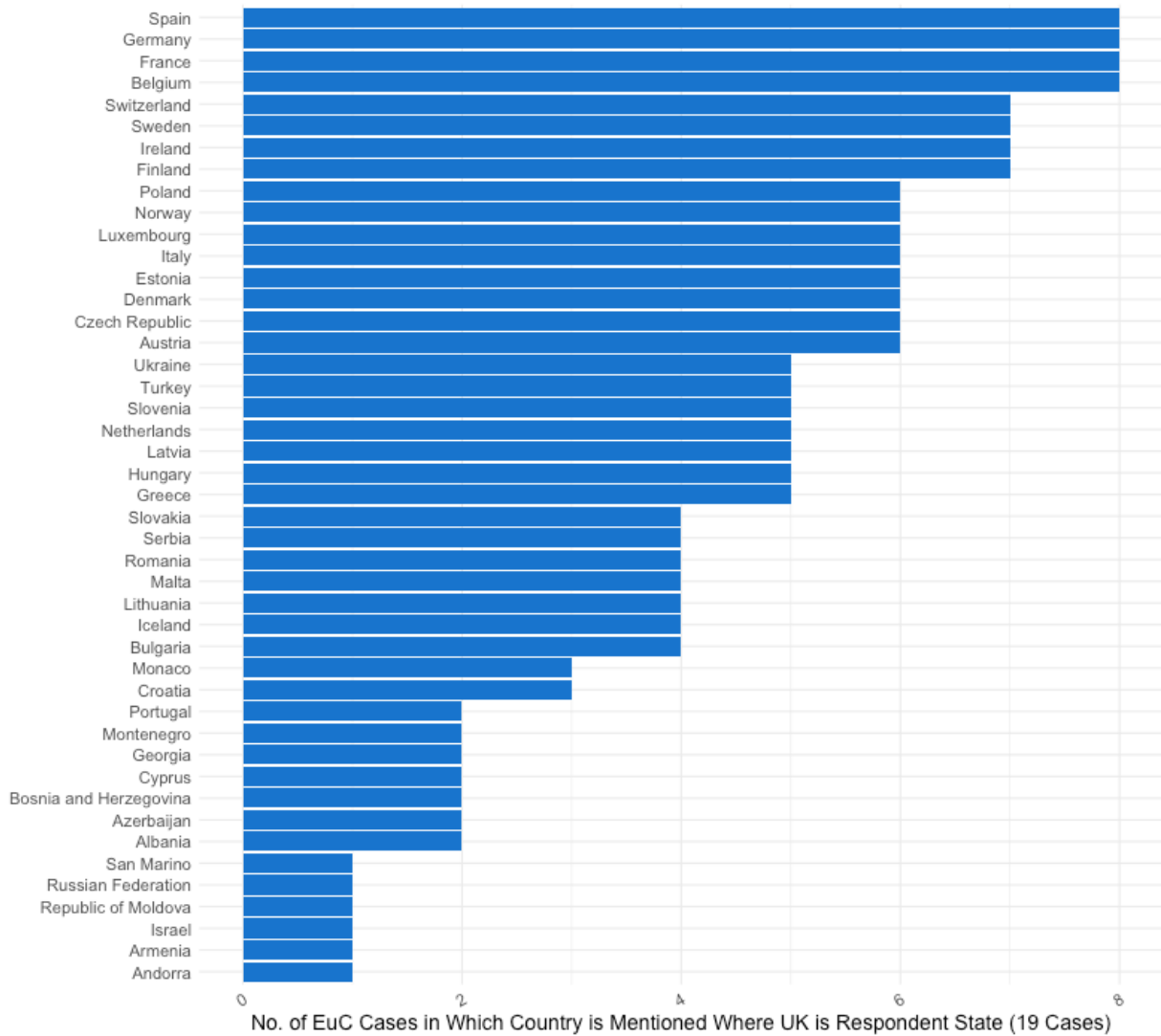
when making consensus decisions.

Figure 5: Number of Consensus Cases With UK as Respondent in Which a Country is Mentioned with Respect to Consensus Language (Hand-Coded Sample)

Discussion and Conclusion

EuC is an increasingly important tool of legal interpretation that is employed by ECtHR judges

when making decisions regarding contentious human rights issues across Europe. Judges take

time and effort to craft legal arguments in hopes that their legal reasoning, not only the outcome

of a particular case, will have an impact on human rights law in the contracting parties across

Europe. Solid legal reasoning, and the use EuC in particular, may confer legitimacy on an ECtHR ruling and make it more likely that actors in national legal systems take the judgment seriously when making decisions at home.

Despite its importance, EuC has proven a difficult legal method to study, to conceptualize, and certainly to quantify. EuC language be identified without a significant amount of qualitative input on the part of highly trained legal experts, partly because it is such a rare event (albeit an extremely important one) and also because it is very context-dependent. We have created a coding scheme that measures three forms of EuC language: Tag 1 (explicit consensus, meaning it uses the term "consensus"), Tag 2 (semi-explicit consensus, meaning it uses terms like "the vast majority of states), and Tag 3 (not explicit, but which contains reference to extraneous sources to suggest EuC logic in the view of legal experts). With this coding scheme, the lawyers have manually coded over 230 Grand Chamber cases. With these coded cases, we can confirm that the Court has increased its usage of EuC language since the 1990s, especially in the last 10 years; we can examine which states are more likely to be on the receiving end of consensus judgments from the Court; and we can examine which states the Court considers when determining whether consensus exists.

However, we have also found that EuC language is hard for a machine learning classifier to identify. There is simply not very much EuC language that a classifier can "learn" from. Moreover, the relative scarcity of EuC language makes it too easy for the classifier to falsely identify paragraphs as containing consensus when they do not. In future research, we hope to be able to use the classifier to guide additional hand-coding of the remaining Grand Chamber judgments. We know that we have high levels of specificity. This means that if the classifier tags almost all instances of consensus language. When working on the uncoded cases, we can be

highly confident that the classifier tags almost all the instances of consensus language. This means that we can focus on those paragraphs and cases that the classifier has identified as containing consensus and attempt to weed out the false positives. We will then be able to provide a complete picture of consensus in the Grand Chamber. Ultimately, we would like to extend our analysis the Chamber cases, as well. However, they are much more numerous and also less likely to contain consensus reasoning, increasing the difficulties we face in using the classifier.

Appendix

Identifying Different Parts of the Cases and When the Court is Speaking

To parse the sections and subsections of the cases, we utilized the structure of the Word version of the cases.[17] For instance, the main sections of the documents are left justified and capitalized, the subsections of the documents are prefaced by roman numerals and are also left justified and capitalized, and the sub-subsections are prefaced by a letter and are bolded and not capitalized. From this, we could identify parts of the document based on the formatting of the text. For each paragraph in the case, we identified whether it was section, subsection, or sub-section. If it was one of these, then we identified the name of the section (or following subsections) and determined whether it was a section of interest. Specifically, these are: The Law Sections and The Relevant Law and Materials Subsection (or sometimes it is a sub-subsection), which appears in either The Facts or The Procedure Sections.

Once restricting the cases to the parts where the Court is laying out its argument, we then had to restrict the documents only to parts where the Court itself is speaking. In these parts of the cases, the Court allows the parties to submit documents and lay out their arguments. While there could be instances of EuC language during these sections, it is not on behalf of the Court, so we need to make sure that we exclude these parts as it could confuse the classifier. Similar to identifying the relevant sections of the cases, the parts where the Court is not speaking are fairly easy to identify through the organization of the document. Any subsection or sub-subsection that contained phrases like "third party", "submitted by the parties", "respondent/intervening

---

[17] Most existing work using quantitative text analysis to analyze ECtHR judgments uses the HTML version of the case as this version, along with its metadata, is easily scrapable from the HUDOC repository.

government", or "the applicants" were determined to be parts where the Court was not speaking, and therefore we removed from the analysis.

To ensure that we were properly removing instances where the Court was not speaking, we gave the lawyers the list of false positive cases tagged by the classifier of our 150 coded cases to review. With this confirmation that we removed instances of the Court not speaking through the organization of the document, we could proceed with processing the rest of the Grand Chamber cases in the same way. Again, as the lawyers manually verify the positively identified cases of the classifier with the out of sample cases, they will also be checking for instances where it is not the Court speaking. When we move to the Chamber, the cases do not have the same structure as in the Grand Chamber cases, meaning that we will not be able to use the same identification process. Instead, we will use a machine learning classifier to identify when the Court is and is not speaking. We have tested the efficacy of this on our already coded cases, and it has high accuracy, suggesting that it will work well out of sample.

Notes on the Machine Learning Classifier

We started our machine learning classifier with our in-sample already coded 150 cases. Prior to running the model, we removed all paragraphs shorter than seven words as well as section names. As instances of EuC language are rare- only occurring in 2.5% of paragraphs (or 342 out of 13,655 in the 150 cases), we oversampled instances of consensus language in our training set. For this, we oversample consensus language and under sample non-consensus language to ensure that there are enough positive instances to train our model. This leads to a ratio of about 1:2 (consensus to non-consensus paragraphs). For each model, which we specify below, we run 10 different random samples for the training and testing set. This allows us to

ensure that specific paragraphs are not driving the results. From these ten different iterations, we then calculate the average F-1, precision, specificity, and recall score for each version of the model. This ensures that we have instances of consensus language in both the training and test sets while also ensuring that specific observations are not driving the results.

We ran multiple different models with different steps of text preprocessing to ensure we had the best model fit. We included four ways of processing the text: one, just lower casing all text and removing special characters and numbers; two, we included the previous steps while we stemmed the text and only selected words that occurred in at least 50 paragraphs; three, we included the steps in the first version while we stemmed the text while only including text that occurred at least 25 paragraphs; and four, we included unigrams and 3-skip-bigrams from stemmed text while removing n-grams that occur less than 25 times or in less than 1% of paragraphs. We used both SVM and Naive Bayes with and without Term Frequency- Inverse Document Frequency (tf-idf) transformation. This in total leads to 12 different combinations of preprocessing and model specifications. Within the in-sample cases, the models that perform the best across the different metrics is the smallest amount of text preprocessing (just lower casing all text and removing special characters and numbers) with SVM and tf-idf transformation.[18]

---

[18] Similarly, the model with the same specifications, except without the tf-idf transformation performs almost equally as well.

References

Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. 2016. "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective." *PeerJ Computer Science* 2: e93.

Cichowski, Rachel A. 2006. "Courts, Rights, and Democratic Participation." *Comparative Political Studies* 39(1): 50–75.

Coppedge, Michael. 2002. "Democracy and Dimensions: Comments on Munck and Verkuilen." *Comparative Political Studies* 35(1): 35–39.

Cox, Gary W., and Mathew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. Cambridge ; New York: Cambridge University Press.

———. 2007. *Legislative Leviathan: Party Government in the House*. 2nd ed. Cambridge ; New York: Cambridge University Press.

Dinas, Elias, and Ezequiel Gonzalez-Ocantos. 2021. "Defending the European Court of Human Rights: Experimental Evidence from Britain." *European Journal of Political Research* 60(2): 397–417.

Dzehtsiarou, Kanstantsin. 2015. *European Consensus and the Legitimacy of the European Court of Human Rights*. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9781139644471 (January 28, 2022).

Epstein, Lee, William M. Landes, and Richard A. Posner. 2012. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Harvard University Press. https://www.degruyter.com/document/doi/10.4159/harvard.9780674067325/html (January 18, 2022).

Finnemore, Martha, and Kathryn Sikkink. 1998. "International Norm Dynamics and Political Change." International Organization 52(4): 887–917.

Gilardi, Fabrizio. 2013. "Transnational Diffusion: Norms, Ideas and Policies." In: Carlsnaes, Walter; Risse, Thomas; Simmons, Beth A. (eds.). *Handbook of international relations.* Thousand Oaks: SAGE, Second Edition, 453-477.

Greer, Steven. 2000. *The Margin of Appreciation: Interpretation and Discretion under the Eur opean Convention on Human Rights*. Strasbourg: Council of Europe Publishing.

Grewal, Sharanbir, and Erik Voeten. 2015. "Are New Democracies Better Human Rights Compliers?" *International Organization* 69(2): 497–518.

Hafner-Burton, Emilie M. 2021. "Elite Decision-Making and International Law: Promises and Perils of the Behavioral Revolution." *AJIL Unbound* 115: 242–47.

Hafner-Burton, Emilie M., D. Alex Hughes, and David G. Victor. 2013. "The Cognitive Revolution and the Political Psychology of Elite Decision Making." *Perspectives on Politics* 11(2): 368–86.

Hafner-Burton, Emilie M., and Kiyoteru Tsutsui. 2005. "Human Rights in a Globalizing World: The Paradox of Empty Promises." *American Journal of Sociology* 110(5): 1373–1411.

Helfer, Laurence R., and Erik Voeten. 2014. "International Courts as Agents of Legal Change: Evidence from LGBT Rights in Europe." *International Organization* 68(1): 77–110.

Hillebrecht, Courtney. 2009. "Rethinking Compliance: The Challenges and Prospects of Measuring Compliance with International Human Rights Tribunals." *Journal of Human Rights Practice* 1(3): 362–79.

———. 2012. "Implementing International Human Rights Law at Home: Domestic Politics and the European Court of Human Rights." *Human Rights Review* 13(3): 279–301.

———. 2014. "The Power of Human Rights Tribunals: Compliance with the European Court of Human Rights and Domestic Policy Change." *European Journal of International Relations* 20(4): 1100–1123.

Howard, Robert M., and Jeffrey A. Segal. 2002. "An Original Look at Originalism." *Law & Society Review* 36(1): 113.

Kapotas, Panos, and Vassilis P. Tzevelekos, eds. 2019. *Building Consensus on European Consensus: Judicial Interpretation of Human Rights in Europe and Beyond*. 1st ed. Cambridge University Press. https://www.cambridge.org/core/product/identifier/9781108564779/type/book (January 24, 2022).

Khosla, Madhav. 2011. "Inclusive Constitutional Comparison: Reflections on India's Sodomy Decision." *American Journal of Comparative Law* 59(4): 909–34.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from

    Political Texts Using Words as Data." *American Political Science Review* 97(2): 311–31.

Lupu, Yonatan, and Erik Voeten. 2012. "Precedent in International Courts: A Network Analysis

    of Case Citations by the European Court of Human Rights." *British Journal of Political*

    Science 42(2): 413–39.

Madsen, Mikael Rask. 2020. "Two-Level Politics and the Backlash against International Courts:

    Evidence from the Politicisation of the European Court of Human Rights." *The British*

    Journal of Politics and International Relations 22(4): 728–38.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov

    Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10(2):

    134–53.

McCubbins, Mathew D., Roger G. Noll, and Barry R. Weingast. 1987. "Administrative

    Procedures as Instruments of Political Control." *Journal of Law, Economics, &*

    Organization 3(2): 243–77.

McCubbins, Matthew D., Roger G. Noll, and Barry R. Weingast. 1989. "Structure and Process,

    Politics and Policy: Administrative Arrangements and the Political Control of Agencies."

    Virginia Law Review 75(2): 431.

Medvedeva, Masha, Michel Vols, and Martijn Wieling. 2020. "Using Machine Learning to

    Predict Decisions of the European Court of Human Rights." *Artificial Intelligence and*

    Law 28(2): 237–66.

Moyer, Laura P., and Susan B. Haire. 2015. "Trailblazers and Those That Followed: Personal Experiences, Gender, and Judicial Empathy: Trailblazers and Those That Followed." *Law & Society Review* 49(3): 665–89.

Munck, Gerardo L., and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices." *Comparative Political Studies* 35(1): 5–34.

Palmer, Barbara. 2001. "Women in the American Judiciary: Their Influence and Impact." *Women & Politics* 23(3): 91–101.

Panke, Diana. 2020. "The European Court of Human Rights under Scrutiny: Explaining Variation in Non-Compliance Judgments." *Comparative European Politics* 18(2): 151–70.

Peat, Daniel C. 2021. "The Tyranny of Choice and the Interpretation of Standards: Why the European Court of Human Rights Uses Consensus." NYU Journal of International Law and Politics 53(2): 381-432.

Reh, Christine, Adrienne Héritier, Edoardo Bressanelli, and Christel Koop. 2013. "The Informal Politics of Legislation: Explaining Secluded Decision Making in the European Union." *Comparative Political Studies* 46(9): 1112–42.

Riker, William H. 1986. *The Art of Political Manipulation*. New Haven: Yale University Press.

Segal, Jeffrey Allan, and Harold J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model*. Cambridge ; New York, NY: Cambridge University Press.

Shepsle, Kenneth A. 1979. "Institutional Arrangements and Equilibrium in Multidimensional

    Voting Models." *American Journal of Political Science* 23(1): 27.

Stiansen, Øyvind. 2019. "Delayed but Not Derailed: Legislative Compliance with European

    Court of Human Rights Judgments." *The International Journal of Human Rights* 23(8):

    1221–47.

———. 2021. "Directing Compliance? Remedial Approach and Compliance with European

    Court of Human Rights Judgments." *British Journal of Political Science* 51(2): 899–907.

Stiansen, Øyvind, and Erik Voeten. 2018. "Backlash and Judicial Restraint: Evidence From the

    European Court of Human Rights." *SSRN Electronic Journal*.

    https://www.ssrn.com/abstract=3166110 (September 26, 2021).

Theilen, Jens T. 2021. *European Consensus between Strategy and Principle*. Nomos

Tomz, Michael. 2008. "Reputation and the Effect of International Law on Preferences and

    Beliefs." Working Paper, Standford, CA: Standford University. Available at:

    https://tomz.people.stanford.edu/sites/g/files/sbiybj4711/f/tomz-intllaw-2008-02-11a.pdf.

    Accessed 2 June 2022.

Voeten, Erik. 2007. "The Politics of International Judicial Appointments: Evidence from the

    European Court of Human Rights." *International Organization* 61(04).

    http://www.journals.cambridge.org/abstract_S0020818307070233 (September 26, 2021).

———. 2008. "The Impartiality of International Judges: Evidence from the European Court of

    Human Rights." *American Political Science Review* 102(4): 417–33.

———. 2021. "Gender and Judging: Evidence from the European Court of Human Rights."

Journal of European Public Policy 28(9): 1453–73.