ABSTRACT

Title of Thesis: COMPUTATIONAL ANALYSIS OF THE CONVERSATIONAL DYNAMICS OF THE UNITED STATES SUPREME COURT

Timothy W. Hawes, Master of Arts, 2009

Thesis directed by: Professor Jimmy Lin
The iSchool
Professor Philip Resnik
Department of Linguistics

The decisions of the United States Supreme Court have far-reaching implications in American life. Using transcripts of Supreme Court oral arguments this work looks at the conversational dynamics of Supreme Court justices and links their conversational interaction with the decisions of the Court and individual justices. While several studies have looked at the relationship between oral arguments and case variables, to our knowledge, none have looked at the relationship between conversational dynamics and case outcomes. Working from this view, we show that the conversation of Supreme Court justices is both predictable and predictive. We aim to show that conversation during Supreme Court cases is patterned, this patterned conversation is associated with case outcomes, and that this association can be used to make predictions about case outcomes.

We present three sets of experiments to accomplish this. The first examines the order of speakers during oral arguments as a patterned sequence, showing that cohesive elements in the discourse, along with references to individuals, provide significant improvements over our "bag-of-words" baseline in identifying speakers in sequence

within a transcript. The second graphically examines the association between speaker *turn-taking* and case outcomes. The results presented with this experiment point to interesting and complex relationships between conversational interaction and case variables, such as justices' votes. The third experiment shows that this relationship can be used in the prediction of case outcomes with accuracy ranging from 62.5% to 76.8% for varying conditions. Finally, we offer recommendations for improved tools for legal researchers interested in the relationship between conversation during oral arguments and case outcomes, and suggestions for how these tools may be applied to more general problems.

COMPUTATIONAL ANALYSIS OF THE CONVERSATIONAL DYNAMICS OF THE UNITED STATES SUPREME COURT

by

Timothy W. Hawes

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Arts
2009

Advisory Committee:

Professor Jimmy Lin, Co-Chair
Professor Philip Resnik, Co-Chair
Professor Wayne McIntosh

Acknowledgments

I couldn't possibly list all the people I want to thank and all the things they have done for me. Please know whether it is listed here or not, I am extremely grateful for everything my friends, family and acquaintances have done for me.

I would like to thank:

Dr. Jimmy Lin and Dr. Philip Resnik, my advisors on this project, for their continually invaluable support, feedback, encouragement and advice not just on this project, but in general.

Dr. Wayne McIntosh and Dr. Michael Evans, for their generosity with their time, opinions and ideas throughout the course of this project. It was a discussion with them that gave initial shape to the conversational view taken in this thesis.

Dr. Amy Weinberg, my first official advisor in the Department of Linguistics, for her excellent guidance and understanding.

Dr. Stephan Greene, for his time and ideas at the earliest stages of this work.

The Department of Linguistics and all of its professors, for their support and guidance.

All of my sources of funding over the past 3 years.

My crack team of proof-readers: Kelly Schultz, Dan Knudsen, Mindy Watson, Mischa Bauermeister, Gordon Freeman, Indira Sriram and Brian Hawes. They noticed more typos than I'd care to admit and each provided excellent suggestions on how to improve my thesis.

All of my friends at the University of Maryland and especially Johannes, Josh, and Greg, for their good humor, support, advice and feedback over the years; Asad for his last minute help saving me hours of highway driving, and also his always enjoyable

conversations; and the many more who should be thanked for everything from invaluable help and support to just being good friends.

All of my friends who have since dispersed across the globe: Dan, Gordon, John, Tim, Mischa, Kevin, Kara and others. You have done more for me than I could ever recount.

I thank Kelly, for her love and support over the years.

And my family: Mom, Dad, Kendra, Tim, Gam, my aunts and uncles (especially Aunt Jane and Uncle Wayne), my cousins (especially John and Kyle) and Chris (who while she isn't technically "family", should be listed here). I appreciate everything you all have done for me.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

The United States Supreme Court plays a significant role in the U.S. Government; the decisions reached by Supreme Court justices have far-reaching implications for the entire American legal system. In this work, we aim to combine conversation analysis with computational techniques in novel approaches for the analysis of the behavior of the U.S. Supreme Court, in terms of both the justices individually and the Court as a whole.

Considerable amounts of work have been done applying computational techniques to the political domain. For example, Mosteller and Wallace (1964) utilized models based on function word counts to identify the authorship of *The Federalist Papers*. Laver et al. (2003) used party manifestos and legislative speeches to identify the ideological positions of political parties in Britain, Ireland and Germany. More directly related to this work is that of Thomas et al. (2006), who examined the content of congressional floor debates and the relationships between congresspersons to determine whether individuals were in support of or opposition to the legislation under discussion. Also, Evans et al. (2007) classified the ideological position of third-party briefs from the briefs' content. We leave further discussion of related work to Chapter 2.

This thesis explores justice *turn-taking* during United States Supreme Court oral arguments and its relationship to other aspects of justice behavior. For our purposes, we will treat each speech segment in the argument transcripts with a single speaker identifier as one *turn*.[1] Thus, the oral arguments are organized into a series of turns produced by the

[1] Due to the Courtroom reporter's handling of factors such as interruption and overlapping speech, this definition of turn is somewhat different from that used in conversation analysis, where they are "turns at talk" composed of units that are grammatically and phonetically realized and "constitute a recognizable

justices and the attorneys before the Court. The first experiments we discuss look at the prediction of the turn-taking behavior of justices by exploring the task of labeling turns with their speakers when this information is unavailable in an oral arguments transcript. The next Chapter is a broad-scale analysis of the turn-taking patterns of justices in various conditions by looking at patterns of when justices typically follow-up on other justices' lines of questioning. Chapter 5 discusses a group of experiments that looks at the turn-taking behavior of justices as a predictor of case outcomes.

This work will be immediately relevant to researchers exploring the behavior of the United States Supreme Court. This view of the conversational dynamics between the justices as both predictable and predictive is one that has received little attention in the literature. By applying computational models to this approach, this work will provide new tools that may be able to open up novel avenues of research for legal scholars. Moreover, this work should also have broader implications. While we have concentrated on applying existing computational tools to a new approach to understanding the Supreme Court, the methods we develop here will be applicable to similar settings where one may wish to link conversational actions to other actions with a real world impact. If this is the case, then these methods will help to provide a deeper understanding of other social institutions and human conversational interaction in general.

While the narrow focus of this work is to produce methods for classification and labeling of the oral arguments of the U.S. Supreme Court, this research was conducted with the broader goal of creating novel approaches for judicial scholars to use in examining the dynamics of the Supreme Court.

---

action in context" (Schegloff 2007; 3-4). Despite this difference, there will still be significant overlap between what we are defining as a turn and what a conversation analyst would define as a turn.

Our primary objective is to gain a clearer understanding of the role of the conversational dynamics of Supreme Court justices. We aim to show that: a) predictable high level patterns exist in the conversational dynamics of the Supreme Court, b) these patterns may be associated with other areas of interest to legal scholars such as voting patterns of the justices, c) this association between linguistic patterns and judicial patterns may be utilized to provide both short term insights (i.e. predicting the outcome of a particular case) and deeper insights about the behavior of the Supreme Court.

In the process of pursuing these objectives we have decided to minimize the need for specialized knowledge and training for feature identification. In order to do this, we minimize theoretical commitments, thus reducing the need for an extensive background in any particular theory of discourse. Moreover, we want to reduce reliance on features that can only be encoded with human judgment and expertise, by favoring features that can be automatically recognized. By restricting ourselves to such conditions we hope to maximize the applicability and reproducibility of our methods, as the reliance on human judgment has hampered both of these qualities in some previous work. Despite this, we expect that higher level information from more sophisticated approaches, such as sentiment analysis, would only add to the value and power of these basic approaches.

Producing *any* positive result for this work is a contribution to the overall understanding of the Court. While small studies using human judgments have produced relatively large positive results, larger studies using automatic methods still achieve relatively small improvements (See Section 2.4). In one case, these automatic methods achieve comparable results to our own work with an order of magnitude more data. Also, when tested on our dataset, these methods achieve considerably lower results. Just as

these larger studies have contributed to the understanding of the relationship between one aspect of oral arguments and case outcomes, positive results in this work should contribute to the understanding of the relationship between conversational interaction and case outcomes. Moreover, given the relative simplicity of our feature sets, the fact that we are able to gain some predictive power at all from these features may be a surprising result for legal scholars (Evans, M. personal correspondence, August 28, 2009).

Thesis Organization

The remainder of this thesis is organized as follows:

- Chapter 2 discusses background on oral arguments, discourse and conversation analysis, computational approaches to discourse and conversation analysis, quantitative research on oral arguments, and the Supreme Court case database used in two of our experiments.
- Chapter 3, Chapter 4 and Chapter 5 cover our three experiment groups dealing with turn sequence labeling, "rose diagrams" of turn-taking and case outcomes and case outcome prediction, respectively.
- The final Chapter offers conclusions from this work and suggests some future research and unanswered questions.

Chapter 2 Background

This chapter contains three main parts. The first part covers the domain knowledge regarding the area of study contained in this thesis, namely, oral arguments and the Supreme Court. The second introduces the linguistic area of study we utilize in this thesis, specifically, conversation and discourse analysis. The third part is an overview of computational studies in discourse analysis as well as a review of both computational and manual studies of the Supreme Court. We include one final section to introduce our source of Supreme Court case data (not including oral argument transcripts).

2.1 Oral Arguments/Supreme Court

As one of the last, and only public, stages a case goes through before the Supreme Court, the importance of oral arguments is often questioned. At this stage, all briefs have been submitted by each side of a case and by *amici curiae*, and the justices have had time to study the details of the case. It is believed that by this time, justices have had sufficient opportunity to make up their mind regarding a case, and so it is often suggested oral arguments play little if any role in justices' decision making process (Rhode & Spaeth 1976; Kurland & Hutchinson 1983; Segal & Spaeth 2002). Kurland and Hutchinson (1983) argue, "There are a few cases in which oral argument serves as a means of discovery by the Justices. But there is no reason why this discovery could not be conducted better by interrogatories than by oral deposition." This view is not just held by academics either: some justices have also expressed these views. Justice Thomas once said, "99 per cent of the time justices have made up their mind when they go to the

bench. Also, there are so many questions you have to elbow your way in" (Rombeck 2002; 5B).

Even for those justices who do view oral arguments as important, it would seem that they do not believe oral augments typically lead a justice to change his or her mind. On the topic of whether oral argument matters, Justice Rehnquist wrote, "I think it does make a difference" though only in "a significant minority of cases….The change is seldom a full one-hundred-and-eighty-degree swing, and I find that it is most likely to occur in cases involving areas of law with which I am least familiar" (Rehnquist 2002). In a 2009 interview, Justice Scalia (who admits that he once believed oral arguments were a "dog and pony show" (Johnson 2004)) said, "A lot of people are under the impression that [oral advocacy] is a dog and pony show. The judges have read the briefs, they come in with their minds made up, and this is just a performance for the benefit of your client. If that's the impression you have, you are just wrong. I have never met a judge who doesn't think that oral argument is important" (Duke Law 2009). However, similar to Rehnquist, he suggested that only in cases where he has not already made up his mind do oral arguments play a role in his decision making.

While the view that oral arguments are unimportant is commonly held, some scholars have also argued against it, suggesting that justices do in fact utilize information gained during oral arguments to make decisions (Johnson 2001, Johnson 2004, Shullman 2004, Johnson et al. 2006). Johnson (2001; 2) points out that up to oral arguments, the majority of information the justices have seen is that which "other actors want them to see and consider", and that justices use oral arguments as an opportunity to get at what they want to "see and consider" in order to make a decision in the case. However, even in

these studies, the strongest conclusion made is that, in typical cases, oral arguments at best are used to refine a justices' opinion, thus having an important impact on the details of a case's outcome but not necessarily on the case's overall outcome.

Johnston et al. (2009a) note David Frederick's observation that oral arguments are composed of conversations between a lawyer, a justice and another "potentially persuadable justice". While the above description of oral arguments should indicate that the existence of "potentially persuadable justices" may be in question, it seems natural to presume that even if justices cannot be persuaded during oral arguments, other justices will still attempt to do so.

2.2 Discourse Analysis

Discourse analysis is a fairly broad subfield of linguistics. Schiffrin et al. (2001; 1) note that discourse analysis is often not strictly defined but usually refers to one of three domains of study; "(1) anything beyond the sentence, (2) language use, and (3) a broader range of social practice that includes nonlinguistic and nonspecific instances of language." Given this broad definition of discourse analysis, it is clear that there is an open view of what exactly is meant by "discourse". Typically, however, the term is used to indicate a language-based communication forming a "unified whole" (referred to as a *text* in the discourse analysis literature), and such communications can take on a variety of forms including written, spoken or signed (Halliday and Hasan 1976, Johnstone 2007). With regard to the domains of study discourse analysis may involve, aspects of this work could fall under each of these categories; while our first experiment looks at (potentially) extra-sentential linguistic units, overall this work is looking at language use in a particular social setting, the Supreme Court, and the relationship between that language

use and the overall behavior of the Supreme Court. As for our particular version of discourse, we are dealing with transcribed spontaneous speech which inherently incorporates both written and spoken language.

Regardless of the form of communication under consideration, three of the key aspects of discourse an analyst is often concerned with are *texture*, *cohesion*, and *coherence*. Texture, the defining characteristic of a text, is identified by Halliday and Hasan (1976; 2) as "the property of being a text…this [texture] is what distinguishes it [a text] from something that is not a text". Take (1) for example.

(1) A: Does the store carry galvanized wire?
B: Yeah, they do.

This simple exchange can be said to have texture, because it can stand alone as (or at least be a part of) a unified conversation.

Contributing to the texture of (1) is the use of reference (anaphora; *they* refers to *the store*) and substitution (*do* stands in for *carry galvanized wire*) in B. Taken together, these lend cohesion to the text, creating texture. Cohesion refers to the relations that exist within a text between separate units in that text and the idea that "the INTERPRETATION of some element in the discourse is dependent on that of another" (Halliday and Hasan 1976; 4). In the example above, in order to interpret B correctly we need A. Cohesion can take on a number of forms, falling under the headings of grammatical cohesion and lexical cohesion. Grammatical cohesion refers to the use of grammatical tools to create cohesive relations in a text; including reference and substitution as in the example above as well as ellipsis (omission of clauses; e.g. Who stole the book? – John ~~*stole the book*~~) and conjunction (linking of clauses; e.g. John went to the bank. *Later* he went to the

movies).[2] We will discuss conjunction more thoroughly later in this Section. Lexical cohesion includes repetition of the same word, or semantically related words such as holonyms (tree-forest), hypernyms (hat-clothing), semantically "close" terms (banana-apple), etc. (Halliday and Hasan 1976, Brown and Yule 1983).

While cohesion deals with overt relations in a text, coherence deals with relations that must be interpreted by an individual listening to or reading a text. Coherent relations are the underlying relations that hold between segments of text (Brown and Yule 1983). Returning to (1) above, while B is a cohesive response to A, we need to appeal to coherence in order to describe it as an appropriate response to A, as cohesion is no guarantee of coherence. For example, suppose we changed B in (1) as we have done in (2). While B is cohesive with A in (2), *they* still refers to *the store*, it is no longer a coherent answer to A.

(2) A: Does the store carry galvanized wire?
B: They are open on Sundays.

Thus, coherence too is a necessary aspect in building an interpretable discourse. For this work, we make the assumption that the texts we are dealing with, as spontaneous conversations between multiple individuals, are in fact coherent discourses at least for the parties involved. And, while it is not necessarily the case across all sorts of text and all relations within a text, we are making the assumption that the majority of cohesive relations existing in the text are representative of underlying coherent relations.

The connection between conjunction and the coherence relations they signal plays a role in Chapter 3. While the collection of potential conjunctive elements in English is extensive, Brown and Yule (1983) offer several examples as summarized in Table 1.

[2] Note that the usage of some terms, such as anaphora and ellipsis, is somewhat different in discourse analysis than in generative linguistics.

Type	Examples
Additive	and, or, furthermore, similarly, in addition,
Adversative	but, however, on the other hand, nevertheless
Causal	so, consequently, for this reason, it follows from this
Temporal	then, after that, an hour later, finally, at last

Table 1 Example conjunctive relation markers (Brown and Yule 1983; 191).

It is important to note that because of the role of cohesion in the interpretation of discourse, these elements do not always identify the relations they are paired with in Table 1, nor are explicit elements required to mark these sorts of relations (Brown and Yule 1983). Nevertheless, overt markers of such relations are abundant in many forms of discourse, and do tend to exhibit some regularity in the relations they identify (as indicated by Table 1), even if the relationship is at times variable.

2.3 Conversation Analysis

Because this work deals with transcripts of oral arguments, it is most closely related to conversation analysis which may be viewed as a branch of discourse analysis.[3] Hutchby and Wooffitt (2008; 13) write the "aim" of conversation analysis (CA in their terms) "is to focus on the production and interpretation of talk-in-interaction as an orderly accomplishment that is oriented by the participants themselves…. CA seeks to uncover the organization of talk…from the perspective of how the participants display for one another their understanding of 'what is going on'". Because of this view, there is a focus on conversation as a sequence of "turns at talk", with each subsequent speaker turn in a conversation indicating the speaker's understanding of the preceding conversation

[3] However, conversation analysis comes with its own tools, methods and procedures for recording and analyzing conversation that we do not make use of. Despite this, many of the topics of interest to the conversation analyst are relevant to this discussion.

(Hutchby and Wooffitt 2008). In the present work we are particularly interested in this sequence of turns, how predictable that sequence is in a setting like the Supreme Court, and the relationship between this sequence and other actions taken by the Court.

The previous discussion of cohesion and coherence can be tied into conversation analysis through a particular aspect of conversational sequence organization known as adjacency pairs. Adjacency pairs include two turns that are usually, but not necessarily, adjacent in conversation, where the first turn "initiates some exchange" and the second turn is "responsive" to the first. These are treated as pairs because not all types of initiations can be followed by all sorts of responses. So while Question/Answer (e.g. **(1**)) and Apology/Acceptance (e.g. (3)) are typical adjacency pairs, Question/Acceptance and Apology/Answer are not (Schegloff 2007; 13-14).

(3) A: Sorry I broke your mug.
B: That's ok.

Regardless of the pair, recognizing a pair as a member of a particular type requires a coherent interpretation of that pair. However, responses to the first part of a pair may include or be entirely composed of elements that are cohesive with the previous turn (4).

(4) A: When are we going to the movies?
B: Later.

Often times, as in the example given, these cohesive elements are conjunctive, linking the first turn to the second with relations related to those in Table 1. For example, if the initiating turn is a statement, a possible response may be to disagree with the statement. In this case, the response may begin with an "adversative" element (5).

(5) A: Let's go to the movies.
B: But I don't want to.

As stated before, this relationship between cohesive elements and coherence relations offers insight into the discussion in Chapter 3.

2.3 Computational Conversational/Discourse Analysis

Though considerable work has been done in the domain of computational discourse analysis, interest in multi-party discourse involving more than two parties is relatively new, instead favoring single and two-party discourse. Broadly speaking, much of computational linguistics that explores language on the document level has focused on single-party discourse, since texts typically represent a single-party discourse. The following is a sampling of representative papers for single, two, and multi-party discourse. We concentrate on a variety of the more popular areas of research in discourse including coherence relation identification and topic segmentation and identification.

For single party discourse (including text and monologue), Mann and Thompson's (1988) Rhetorical Structure Theory (RST) has been used as a framework for identifying coherence relations in texts from a single author (Marcu, 1997; Corston-Oliver, 1998). Marcu and Echihabi (2002) developed an approach to automatically identify discourse relations that hold between sentences and within sentence parts from a very large corpus of unannotated sentences drawn from textual resources. Grosz and Hirschberg (1992) used a Classification and Regression Tree analysis to identify discourse segments (building on the theory of discourse discussed in Grosz and Sidner (1986)) in Associated Press articles read aloud by news broadcasters. Morris and Hirst (1991) explored "Lexical Chains" (spans of related words in a discourse; in this case text) as a means for modeling lexical cohesion.

In the area of two-party dialog, Stolcke et al. (2000) modeled "dialogue acts" in telephone conversations for automatic labeling.[4] Forbes-Riley and Litman (2004) used acoustic and non-acoustic cues in spoken dialogs to predict the emotional state of students in one-on-one interaction with tutors via AdaBoost with decision trees. Gurevych and Strube (2004) used (manually disambiguated) noun senses from WordNet to summarize the content of telephone-based conversations. Finally, Williams and Young (2007) developed an approach for managing spoken human-machine dialogue.

Much of the existing research on conversation involving three or more parties has been conducted using the International Computer Science Institute (ICSI) meeting corpus (Janin et al. 2003), though other corpora are available (e.g. TalkBank, which includes U.S. Supreme Court oral arguments as a subset of its documents (MacWhinney et al. 2007)). Galley et al. (2003) use a lexical cohesion approach to create an unsupervised method of topic segmentation in multi-party ICSI meetings, while Purver et al. (2006) offer an unsupervised method for topic segmentation and identification using Bayesian inference. Galley et al. (2004) used lexical, contextual and durational cues to identify agreement and disagreement between speakers turns in ICSI meetings.

2.4 Quantitative Oral Arguments Research

To date, there have been several studies dealing with Supreme Court oral arguments. Johnson et al. (2009b) examine factors that may be involved in determining why and when justices will give a dissent from the bench, including the number of questions asked by the Court during oral arguments. This study found a small effect in the relationship between dissents from the bench and case activity measured by the

[4] Dialog acts are often one part of an adjacency pair, e.g. "STATEMENT, QUESTION,… AGREEMENT, DISAGREEMENT, and APOLOGY" (Stolcke et al. 2000).

number of questions asked during oral arguments. In work related to our first experiments, Yuan and Liberman (2008) conducted speaker identification experiments using audio transcripts of oral arguments from 78 cases from the 2001 term.[5] For the 800 "clean" test samples used, 98% speaker identification accuracy was achieved by training 8 justice specific speech recognition models, applying each model to a test utterance, and using the model with the highest score to identify the justice.

We will now discuss several studies aimed at forecasting case outcomes, which are summarized in Table 2. Wrightsman (2008) details several attempts to use manual quantitative and qualitative analysis to predict votes. The first of these examples recounts *New York Times* Supreme Court reporter Linda Greenhouse's prediction of case outcomes based solely on oral arguments using her experience as a courtroom reporter. Of 27 articles she prepared based on oral arguments 17 contained predictions, 12 of which were correct (and one held-out because the case was dismissed). The second example is an analysis of 28 cases from the 1980 and 2003 terms by John Roberts. By determining which side was asked the most questions he was able to determine the winner in 24 of those 28 cases studied. The third is a study by law student Sarah Shullman, who attended 10 argument sessions, and recorded information about each question asked including the content, the speaker, the level of "hostility", and the tone of the speaker's voice. After analyzing 7 cases, Shullman also settled on a "most questions asked" rule that predicted the winner in 6 of the 7 cases analyzed and the 3 held out cases. However, as Wrightsman (2008; 133) notes, "determining what constitutes a 'question' is not so simple". For example, Wrightsman (2008; 136) writes, "interaction

[5] Audio transcripts were accompanied by written transcripts, speaker identifications and manual word-alignment from the OYEZ project (http://www.oyez.org/) (Yuan and Liberman 2008).

between advocates and justices do not follow in a discrete manner; two justices may begin to speak at the same time, a justice may interrupt an advocate, and justices may make elongated statements that may contain several questions." From an even more basic standpoint, it is not clear whether or not researchers limit questions to interrogative statements. Without explicitly identifying how questions are to be counted, replicability of these sorts of experiments will be inherently shaky.

Study	Cases	Accuracy	Method	Manual
Greenhouse	16	75.0%	Experience	yes
Roberts	28	85.7%	Most Questions Asked	yes
Shullman	10	90.0%	Most Questions Asked	yes
Wrightsman	24	42%	Most Questions Asked	yes
Ruger et al.	68	75%	Case metadata	no
Johnson et al.	~2000	66.2%/67.5%	Most Questions Asked / Words Used	no

Table 2 Summary of previous studies. 'Manual' indicates whether or not the study used manual methods of outcome forecasting (the alternative being automatic methods). 'Cases' indicates the number of cases tested in the study.

The final study discussed in Wrightsman (2008) was conducted by Wrightsman and a student. It examined 24 cases from the October 2004 term, 12 of which were identified as "very ideological" and 12 of which were identified as "definitely not-ideological". For each of these cases they determined whether each justice's "overall pattern of questions" was "unsympathetic" to a particular side in the case, as well as the number of questions asked of each side. While no definition of "unsympathetic questioning" is provided, they do provide an example of an unsympathetic statement from *Small v. United States*: in arguing for the side of Small, Justice O'Connor said, "Congress thinks about the United States, our country, and if it means to say something will take place in other places in the world, it says so clearly".

While they do not report absolute accuracy values for the "unsympathetic" questioning approach, they do point out that 87% of the unsympathetic comments were directed at the losing side in the ideological cases and 69% of the unsympathetic comments were directed at the losing side in non-ideological cases.[6] Perhaps more importantly, they report that the "more questions asked" rule employed by Shullman and Roberts led to 42% accuracy. In an attempt to rectify the discrepancy for the "most questions asked" rule, results remained mixed, though a potential pattern emerged; namely this rule seems to be most useful in ideological cases and least useful in non-ideological cases.

While there has been extensive quantitative study on Supreme Court forecasting, computational work has been rather limited with only two studies (Ruger et al. 2002, 2004 and Johnson et al. 2009a). Ruger et al. (2002, 2004) utilized classification trees built from 6 metadata features for 8 years' worth of Supreme Court cases under Rehnquist (658 cases). The metadata used include:

> (1) the circuit of origin for the case; (2) the issue area of the case, coded from the petitioner's brief using Spaeth's protocol; (3) the type of petitioner (e.g., the United States, an injured person, an employer); (4) the type of respondent; (5) the ideological direction of the lower court ruling, also coded from the petitioner's brief using Spaeth's protocol; and (6) whether or not the petitioner argued the constitutionality of a law or practice.
>
> (Ruger et al. 2004)

[6] Though presumably not the case, their method of reporting leaves open the extreme possibility that only two cases contained unsympathetic questioning and for those two cases 87% and 69% of the unsympathetic questions were directed at the losing side. Of course, if this possibility is open, less extreme scenarios about the distribution of the questions are possible. In any case, this does not give a clear picture of the accuracy provided by this approach.

The authors argued that each of these features could be identified by a non-expert, and indeed all but the 6th feature can be found in the Spaeth database (Spaeth 2009).

They used the classification trees to predict cases for the 2002 term prior to the case's decision (68 cases). Finally, results from their classification trees were compared to those of legal experts including "71 academics and 12 appellate attorneys", each of whom have "written and taught about, practiced before, and/or clerked at the Supreme Court". The model performed with an absolute accuracy of 75%, while experts performed at only 58.8% (with results for 10.3% of cases "inconclusive"). Not reported for this timeframe is the proportion of cases decided in favor of the petitioner or respondent. However, based on the term they report using and the cases they held out, it appears that the Court reversed 69.1% of cases during this period.[7]

A more recent and much more comprehensive study was conducted by Johnson et al. (2009a). This study examines all cases from 1979 to 1995 ("over 2000 hours"), testing the "most questions asked" hypothesis. Two logistic regression models are created in this study, the first utilizing the difference in number of questions asked of each side, and the second utilizing the difference in number of words used to discuss the case for each side. In addition to these two main features, features are included in each model to control for potentially confounding factors. These include a "measure of the ideology of the median justice on the Court", the direction of the lower court's decision, a variable to code the interaction of these two previous variables, two variables to code if the Solicitor General participated as amicus curiae on behalf of the petitioner and the respondent and two variables indicating whether amicus briefs were submitted on behalf of the petitioner

[7] Note that there is generally a reversal bias, but that this varies over time. 69.1%, however, is somewhat higher than the typical rate of reversal which is closer to 64%-66%.

and/or on behalf of the respondent. While each of the "questions used" and "words used" variables were the least informative variables in each of the models, they report small, but noticeable effects for these two models with, 66.2% accuracy for the question difference model and 67.5% accuracy for the word difference model.

While the results show relatively low accuracy, given that the Court's tendency to reverse cases is around 64%, they do provide information to suggest that in extreme cases (>2 standard deviations from the mean difference in questions asked) the probability of a case being affirmed ranges between 18% and 39%. They report similar correlations with the distribution of the difference in words used for each side. Thus these results do suggest that despite the conflicting results presented by Wrightsman (2008), there is in fact some relevance to the "most questions asked" hypothesis (and more generally, a "more attention given" hypothesis). However, as is discussed in Chapter 5, we find that for our own data set, the "most questions asked" rule is not predictive across the corpus, though, as suggested by Johnston et al. (2009a) it does provide some benefit in the extreme cases.

Though not explicitly a forecasting study, the work of Johnson et al. (2006, 2007) is also closely related to this work. They used Justice Blackmun's records of the quality of arguments by individuals before the Court to examine the relationship between quality of oral arguments and case outcomes. In addition to Justice Blackmun's records, they attempted to determine if any other factors such as attorney background and justice and attorney policy preferences had an impact on the quality of arguments presented to the Court. Their findings suggest that when the quality of one side's oral arguments are significantly better than another's, the case is more likely to go to the side with the higher

quality arguments, and that an attorney's background may be helpful in determining the quality of arguments they will present. This advantage is as high as a 77.9% chance of reverse when the petitioner's arguments are "manifestly better" than the respondents, and as low as 34.9% chance of reverse in the converse situation.

2.5 Spaeth Supreme Court Database

Much of the work in this thesis utilized the Spaeth Supreme Court Database (Spaeth 2009; henceforth Spaeth database). The Spaeth database is a comprehensive listing of Supreme Court cases and accompanying variables dealing with the "background" of the case (e.g. the origin of the case, the parties involved in the case, the issue area), "chronological variables" including important dates of the case, the identity of the chief justice and the natural court, "substantive variables" such as the issue area of the case and the direction of the decision, "outcome variables" including the winner of the case, and "voting and opinion variables" identifying the votes and opinions issued in the case.

Often cases can involve multiple legal provisions or issues. In these instances, multiple listings are provided for each case. These listings separate variables that would otherwise be conflated. As suggested in Benesh (2002) we concentrate on the "case citation" listing as we "[want] to study decisions in the aggregate and [want] to count each decision only once."

Chapter 3 Sequence Labeling[8]

The work contained in this section aims to address our first objective; to demonstrate that conversational patterns exist in Supreme Court oral arguments. This is accomplished by constructing a sequence labeling task that identifies speakers from turn content. Given a sequence labeling task, if speakers can be identified from the content of the turns *and* increasing the turn history in a model for sequence labeling improves performance, it indicates that patterns exist in the turn-taking behavior of Supreme Court justices.

In a typical labeling task the objective is to identify present, but unobservable information (hidden variables) from observable information (observed variables). An example of a common sequence labeling task is part-of-speech (POS) tagging. In POS tagging, the objective is to identify the parts-of-speech (e.g. noun, adjective, preposition, determiner, conjunction, etc.) for words in a sentence. Framed as a sequence labeling problem, the hidden variables are the POS of each word and, in the simplest case, the observed variables are the words. Because the same words in different sequences may have different POS, one usually wants to make use not only of the words themselves, but of sequential information as well, such as the order of words or the sequence of the predicted POSs. Because of this, POS tagging is often approached with graph based statistical models that can easily make use both of the features in a sequence (i.e. words) and the sequence itself (e.g. DeRose 1988, Lafferty et al. 2001, Toutanova et al. 2003).

[8] This work was originally published in Hawes et al. (2009). Figures in the following Sections are from this paper. Other discussion will either closely coincide with or match the content of this paper. Discussion is expanded and details are included to highlight the relevance of this work to this thesis.

Similar to POS tagging, we can construct a task where the observable information is a sequence of turns, and the hidden variables are the identities of the speaker for each turn. Supreme Court transcripts prior to 2004 offer an immediately relevant example, as justices were not uniquely identified for these cases.

3.1 Methods

Data Preparation

Though the cases used for each experiment set vary, all experiments share a common data preparation approach. Transcripts of oral arguments are posted the same day a case is argued in PDF format. Transcription is conducted by the Courtroom reporter, Alderson Reporting Company. While details of the transcription process are not given, the character and infrequency of errors would indicate that transcripts are created manually.[9] For each segment of speech by a single speaker, transcripts contain the speaker's name (i.e. Speaker ID) and the content of the speech segment. For all experiments, each segment is treated as one speaker *turn* and thus the transcript is treated as an approximation of the turn sequence during the entire case.[10] Finally, transcripts contain several non-content items including opening and closing time stamps and headers for the oral and rebuttal arguments of each litigant (Table 3).

[9] For example, typos in speaker IDs (i.e. non-content text) such as JUSTICE KENNY instead of JUSTICE KENNEDY, or JUDGE ALITO instead of JUSTICE ALITO.

[10] Of course, this sequence can only be an approximation; there is no duration information, only coarse overlap information, and other discourse information such as fillers (i.e. um) are often disregarded.

Symbol	Examples
TIME	(11:08 a.m.), (Whereupon, at 12:08 p.m., the case in the above-entitled matter was submitted.)
START-ORAL	ORAL ARGUMENT OF JEAN-CLAUDE ANDRE ON BEHALF OF THE PETITIONER, ORAL ARGUMENT OF KANNON SHANMUGAM ON BEHALF OF THE RESPONDENTS
START-REBUTTAL	REBUTTAL ARGUMENT OF JEAN-CLAUDE ANDRE ON BEHALF OF THE PETITIONER

Table 3 Examples of non-content items from the transcript of the oral arguments from *Ali v. Federal Bureau of Prisons* (06-9130) with the special symbols used to identify these items in our experiments.

All transcript PDFs were converted to XML format using an off the shelf utility, followed by custom built automatic cleanup to remove extraneous formatting. Cleanup code and cleaned transcripts will be made available at http://www.umiacs.umd.edu/~twhawes/oralarguments/index.html.

Corpus Description

At the beginning of this study the Court's 2007 term had not yet completed, and prior to the 2004 term justices did not have unique speaker IDs. Thus we limited the corpus to the 2004-2006 terms. For the sake of consistency, we also filtered out cases that followed an atypical format.[11] For example, those cases that included arguments from amici curiae.

Feature extraction

From the XML formatted cases we extracted the case content including: speaker IDs, speaker turn content and non-content items in the transcript. Turns were extracted as speaker ID/content pairs. From the content of each turn, we extracted features as shown in the Features Section (c.f. Figure 3).

[11] Filtered out cases include: 02-1472, 04-1067, 04-473b (*Garcetti v. Ceballos (Reargued)*), 04-944, 05-1342, 05-1575, 05-204, 05-705, 05-746, 05-9222, 06-484, 06-5247, 06-5306, 06-593, 105 Orig. (Kansas v. Colorado) and 128 Orig. (Alaska v. United States).

Labeling

We extract from each unit x a set $\bar{x}$ of features, and our models predict the labels y_i for a sequence, yielding $\{(\bar{x}_1, y_1), \ldots, (\bar{x}_n, y_n)\}$. The labels y_i comprise a set of 15 symbols: 11 for the justices (one for each), one to represent the lawyers (either on behalf of the petitioner or respondent), plus one special symbol for time stamps and two additional special symbols to encode the section headings (i.e. START-ORAL and START-REBUTTAL).

Figure 1 shows the frequency with which each of the justices spoke across all cases in the corpus. Not included are the non-justice parties from each side, who produce 47.4% of all turns. Also not included are the special symbols, which comprise 2.2% of symbols in the corpus. While the Court is only composed of 9 justices at any given time, we report 11 in Figure 1 due to changes in court membership, including Robert's replacement of Rehnquist and Alito's replacement of O'Connor. Because these justices do not span this entire corpus, their empirical probability should be lower than that of the justices' true tendency to speak during oral arguments (this, in turn, has an impact on our experimental results).

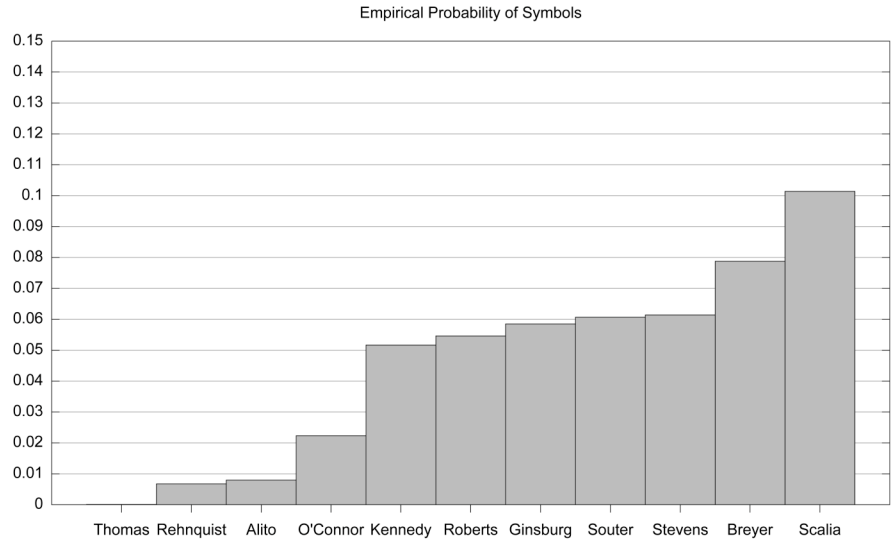


Figure 1 Empirical probability of each justice symbol in the corpus (Hawes et al. 2009).

Because we are predicting sequential labels from a collection of features, conditional random fields (CRFs; Lafferty, McCallum, & Pereira, 2001) are a straightforward choice for this task. CRFs utilize undirected graphs to model the conditional probability of an unobserved sequence of labels (*Y)* given some observable sequence of features (*X)*. CRFs are preferable to Hidden Markov Models (HMMs) in many sequence-labeling tasks because they relax stringent conditional independence assumptions made by generative models. CRFs have been empirically shown to work well for a variety of text processing tasks, including POS tagging (Lafferty et al. 2001), shallow parsing (Sha & Pereira, 2003), and named-entity recognition in the biomedical domain (Settles, 2004). Although the underlying structure of a CRF can take a variety of forms, a linear chain of labels (Figure 2) is often assumed for sequence-labeling tasks because they allow for efficient inference and decoding using the forward-backward and Viterbi algorithms (Sutton and McCallum 2006). Figure 2 corresponds to a first-order CRF, which determines probabilities using features at the current label along with the

previous label; similarly, a second-order CRF corresponds to a model that determines probabilities using features at the current label along with the previous two labels. For this work we used the MALLET implementation of CRFs (http://mallet.cs.umass.edu).

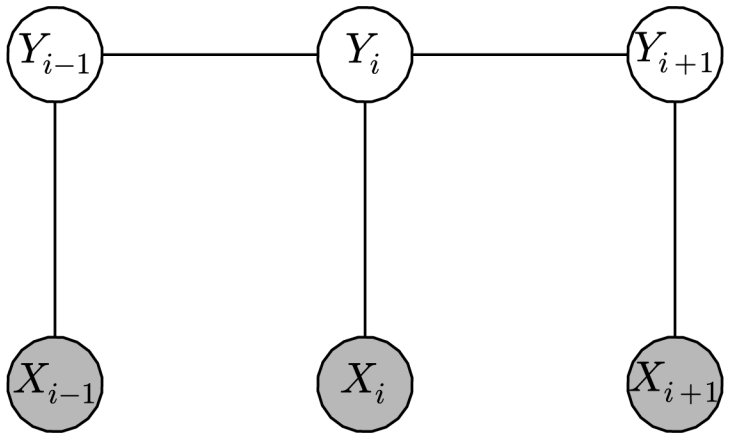


Figure 2 Diagram of a linear chain of labels, where X_i is a group of observed features and Y_i is a label

Features

The following is a discussion of the features used for this task. Note that an additional, contentless feature (*T*) was also used for every turn in order to ensure that all turns had at least one feature in the sequence.

Unigrams

Unique tokens, white space and punctuation separated, were extracted from each turn, ignoring stop-words. One feature for each token used in a particular turn was included in the feature set for that turn indicating the presence of that token. By including unigrams in our feature set, we are essentially creating a "bag-of-words" language model.

Because this is among the simplest possible approaches for this task, we treat unigrams as our baseline feature set.

Discourse Markers (DM)

All interpretable discourse is composed of discourse relations, which serve to connect each unit of discourse. Correct interpretation of these relations is necessary in order to correctly interpret a discourse. Because we can safely assume that oral arguments are an interpretable discourse (at least for all parties involved) we can infer the presence of these coherence relations, not only between an individual speaker's utterance but between the utterances of separate speakers. Instead of attempting to identify all of these relations automatically, however, we instead rely on discourse markers, which have traditionally been viewed as overt cues for underlying discourse relations (cf. conjunctive cohesive elements, Section 2.2).

Both semantically and syntactically optional, discourse markers are typically viewed as pragmatic units used to link clauses in a discourse (Schiffrin 1987). As overt cues of discourse relations, discourse markers are a prime example of conjunctive cohesive elements of a discourse. For this task, we compiled a list of approximately 700 potential discourse markers identified through manual examination of the corpus and in the literature (Marcu, 1997; Oates, 2001).[12] Finally, we make the simplifying assumption that any turn initial string that matches a member of this list is a discourse marker; a condition met in approximately 50% of turns. If multiple adjacent discourse markers appear at the beginning of the string, all were included. Consider an example from

[12] Manual examination of the corpus may be seen as viewing test data prior to testing. The author readily admits this list would have ideally been compiled from out-of-sample documents. However, note that the task is to examine the impact of discourse markers, not to identify discourse markers. Because all potential discourse markers were included using this method, we view this as parallel to annotations in the test data for a task that requires such information.

Kansas v. Marsh (Reargued) (2006): "JUSTICE BREYER: *Okay, well*, what do you say to –", from which we extract two discourse markers (italicized). Because the discourse marker list is composed of both single and multi-word discourse markers, and because the majority of single word discourse markers are also stop-words, there is very little overlap between the *Unigram* feature set and the *DM* feature set.

Personal Reference (Ref)

Finally, we included a feature set for references to individuals. This feature set included four types of features: justice's names, honorifics (i.e. "Your Honor"), second person pronouns, a single feature for any justices mentioned, and a single feature for every non-justice name.[13] Instances of these features were identified using simple pattern matching, which we found to be sufficient for most instances of address due to the formal nature of Supreme Court discourse. Thus, this works well as a basic model for direct address closely related to that discussed in Jovanovic and Akker (2004).

However, one should note that as a consequence of using simple pattern matching and no additional or more sophisticated approaches, all instances of reference are included regardless of the referent. While a subset of these references include direct references to an individual who either spoke or will speak in adjacent turns, the direct address feature set also includes references to individuals present, but not currently participating in the discourse, and to individuals who are not participating in the discourse at all. While each of these different classes of "individual mention" make a distinct contribution, each contribution made is potentially useful in modeling the

[13] The second to last feature was included to account for highly variable mentions of justices that were not serving on the Supreme Court during the case. A single feature was used in this final case because of the high variability across cases of non-justice names. Note however, that the majority of these latter namings within a case typically refer to the party currently presenting oral arguments or other individuals involved in the case.

conversational dynamics of the Court. Because references are typically made to someone who recently spoke or will speak (because they have been addressed), for each turn we include the reference features from the immediately adjacent turns but not the current turn. Approximately 40% of turns contained at least one instance of personal reference. Finally, as with discourse markers, because unigrams are filtered for stop-words and contain only single tokens there was little overlap between the direct address features and the unigram features. Figure 3 provides an example of the features extracted from a sequence of turns.

Turns from *S. D. Warren Co. v. Maine Bd. of Environmental Protection* (04-1527)

JUSTICE SOUTER: -- "reinforcing," and maybe it's "changing." I mean, you're characterizing it one way. We start with a different canon of meaning, and that is that we look to the words around which, in connection with which, the word is used. In here, it's being used without certain modifiers or descriptive conditions. In other cases, it is being used with them. And that's a good reason to think that probably the word is intended to mean something different in those situations.

MR. KAYATTA: Well, I would -- I would hesitate, Justice Souter, to go from taking a specific word, like "discharge," and, therefore, saying that it meant something that is both more general and much more easily set.

JUSTICE SOUTER: No, but your argument, I thought, was simply this, that it uses "discharge" in, you know, X number -- I forget how many you had -- and it's perfectly clear that in most of those instances it requires an addition; and, therefore, it should be construed as requiring it here. My point was that in a great many of those instances, the statute is not merely using the word in isolation; it's using it in connection with a couple of other words, like "discharge a pollutant." And it, therefore, number one, makes sense to construe "discharge of a pollutant" differently from "discharge." That's the -- that's the only point.

Features

Souter 1:

Unigrams: *cases, word, start, changing, connection, words, modifiers, meaning, reinforcing, reason, situations, intended, characterizing, good, canon, descriptive, conditions*

Discourse Markers: -

Direct Address: *you*

Kayatta 1:

Unigrams: *meant, discharge, word, set, justice, souter, easily, taking, specific, general, hesitate*

Discourse Markers: *well*

Direct Address: *Justice_Souter, JUSTICE*

Souter 2:

Unigrams: *argument, simply, requires, sense, discharge, construe, clear, thought, construed, point, number, great, word, connection, requiring, forget, words, couple, addition, differently, perfectly, statute, instances, isolation, pollutant, makes*

Discourse Markers: *no, but*

Direct Address: *your, you*

Figure 3 Example of features extracted from a transcript segment.

3.2 Experiments

For our experiments we utilized four combinations of features:

- Unigrams (Unigrams)
- Unigrams plus Discourse Marker Features (Unigrams + DM)
- Unigrams plus Personal Reference Features (Unigrams + Ref)
- Unigrams plus Discourse Markers plus Personal Reference (Unigrams + DM + REF)

With these features we conducted sequence prediction using both first and second order CRFs.

All experiments were evaluated using *k*-fold cross validation. *k*-fold cross validation is a common evaluation technique wherein data is segmented into any number (i.e. *k*) of non-overlapping subsets of instances, or *folds,* where *k* is less than or equal to the number of individual instances in the data set. For each subset s_i of the *k* subsets, a model is trained on the other *k*-1 subsets, and then evaluated using s_i as a test set. Finally, results from each iteration of testing are combined, typically through averaging (as in our experiments). We used 10-fold cross validation to evaluate our first-order models and 2-fold cross validation to evaluate our second-order models.[14]

Results

Results are reported as the F-score for sequence prediction. F-score is the harmonic mean of precision and recall. We used an equally weighted F-score as the simplest measure of precision and recall. Figure 4 shows the 10-fold cross validation

[14] The choice to use 2-fold cross validation for second order models was based on the significantly longer training time for this order of CRF as compared to first order CRFs.

results using first order CRFs. We report only those justices who regularly spoke in cases during their time on the bench and no other symbols.[15] Each justice category has been annotated with the relative improvement from *Unigrams* to the *Unigrams + DM + Ref* condition.

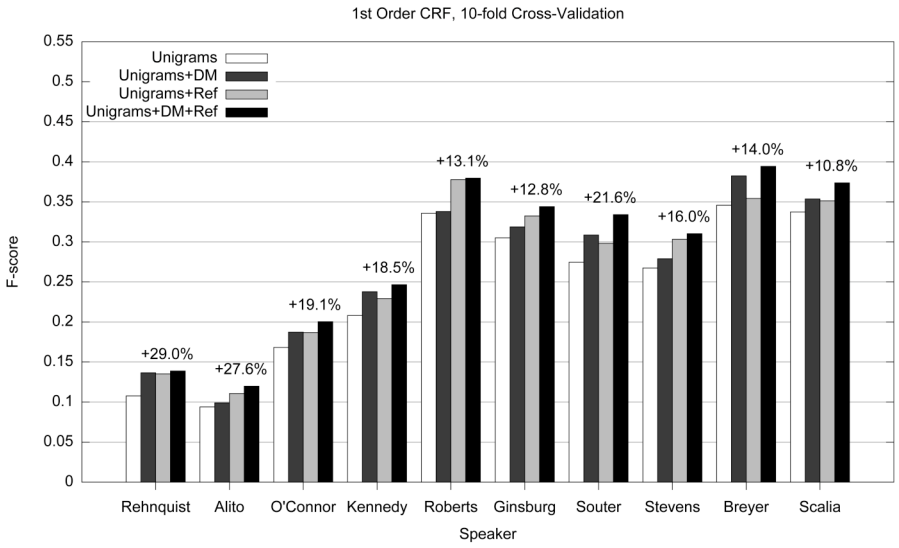


Figure 4 1st CRF 10-fold Cross-Validation Results. Annotations represent the relative improvement over *Unigram* baseline for the *Unigram + DM +Ref* condition (Hawes et al. 2009)

For the *Unigrams + DM* and *Unigrams + Ref* conditions we see relative improvement over *Unigrams* for all justices; however, there is variability across justices as to which of the two provides the greatest relative improvement. The use of both personal reference and discourse markers, in addition to unigrams, provides greater relative improvement than all other conditions for each justice.

Figure 5 shows the 2-fold cross validation results for second order CRFs. As with the first order graphs, justice categories have been annotated with relative improvement

[15] Thus we do not report section headers, the TIME symbol, the L symbol or Thomas (who spoke too infrequently to model).

from the *Unigram* condition to the *Unigram + DM + Ref* condition. For all justices but Alito and Rehnquist we see a relative improvement in all conditions as well as a similar pattern across conditions within justices. The decrease in performance for Alito and Rehnquist is to be expected given that these two justices cover the smallest portions of the corpus compared to all other justices who speak regularly. Because of this, sequences with their symbols appear infrequently across the corpus, and so will either be less evenly distributed throughout cross-validation folds or contain less training data per fold. The overall increase in F-score for all other justices (as compared to Figure 4) in all conditions indicates that increasing speaker history is, as expected, beneficial in modeling justice turn-taking behavior. It would appear that the second-order CRF allows us to capture both complex interactions between justices as well as individual justices' tendency to continue speaking to a lawyer without interruption from other justices.

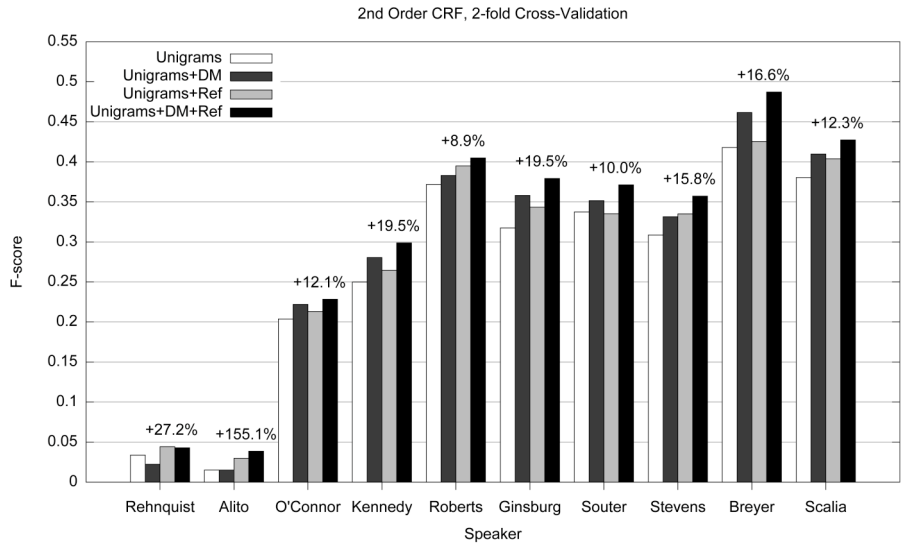


Figure 5 2nd order CRF 2-fold Cross-Validation Results. Annotations represent the relative improvement over *Unigram* baseline for the *Unigram + DM +Ref* condition (Hawes et al. 2009)

Figure 6 contains the overall accuracy for both first and second order CRFs in each condition, where accuracy is simply the proportion of correct predictions to the total number of predictions. Each bar has been annotated with its relative improvement over unigrams for their respective model orders. Error bars were calculated as the 95% confidence interval as computed by the Clopper-Pearson method for inferring exact binomial confidence intervals (Clopper & Pearson, 1934). The confidence intervals indicate that for both first and second order models, the inclusion of discourse markers or personal reference features provides a significant improvement over unigrams alone, though these two conditions are not significantly different from each other. However, the inclusion of both feature sets does provide a significant improvement over both of these conditions for both first and second order models.

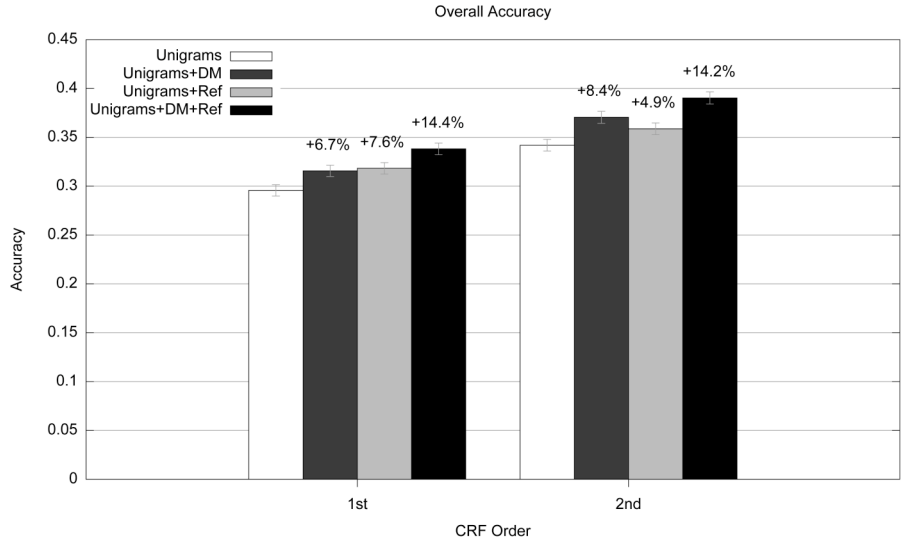


Figure 6 Overall accuracy of first and second order CRFs. Bars are annotated with the relative improvement over *Unigram* baseline. Error bars are the 95% confidence interval as calculated by the Clopper-Pearson method for inferring exact binomial confidence intervals.

Discussion

Interestingly, these results show that the inclusion of features such as discourse markers and instances of personal reference do add information that help in identifying who was speaking when in a discourse. While the results are considerably lower than the acoustic approach to speaker identification of Yuan and Liberman (2008) it should be noted that while our tasks are related, they are also distinct. Their work focuses on the use of acoustic differences in individuals' speech and how this can be applied to speaker identification in acoustically complex environments. In contrast, our work aims to understand the turn-taking patterns of justices in the Supreme Court through the relationship between turn content and turn organization, and we use speaker identification as a task to gauge our progress towards this goal. These results provide significant improvement over a unigram baseline model, and we see significant improvement from first order models to second order models. This indicates the existence of high-level patterns in justice turn organization during Supreme Court oral arguments.

Though we are looking for positive results with our work, we are also looking for tools to help legal scholars. How then, might these results or this work in general be used as such? The fact that we have identified predictable patterns in turn-taking may be of interest to legal scholars. Though they may have had such an intuition about the Court (perhaps, noting that there is a pecking order amongst the justices, with the chief justice at the top, followed by the other justices organized by seniority), these results make this fact explicit. Additionally, the work presented here is a novel approach for understanding the Supreme Court. By utilizing these methods, legal scholars will have new tools for addressing questions about the Supreme Court, and a variety of new questions.

Chapter 4 Visualizing Dynamics

This chapter addresses the second goal of this thesis, to demonstrate that the patterns indicated in the previous chapter can be associated with case outcomes. To accomplish this, we explore the relationship between turn-taking patterns during oral arguments and case outcomes, via a multi-dimensional charting technique. We created charts for sets of cases belonging to a variety of outcomes and case conditions, and examine the relationship between justices' voting record and their turn-taking behavior in these conditions. By comparing these charts we create a picture of the relationship between the voting and conversational behavior of justices.

In this Chapter, as well as the next, we deal with justices' ideology. This is often discussed throughout the media, and often held as common knowledge. However, there have been a number of studies quantitatively examining the ideology of justices. For example, Martin-Quinn scores estimate the "ideal point" (i.e. a point on an attitudinal scale, in this case ideology) for each justice (Martin and Quinn 2002). Martin-Quinn scores are regularly published at http://mqscores.wustl.edu/measures.php. On the Martin-Quinn scale, negative numbers indicate a liberal ideology while positive numbers indicate a conservative ideology. Table 4 summarizes the mean Martin-Quinn score for the justices for the three years covered in our selection of cases.

Justice	Martin-Quinn score
Thomas	4.37
Scalia	2.75
Alito	1.63
Roberts	1.6
Kennedy	0.41
Breyer	-1.41
Souter	-1.51
Ginsburg	-1.54
Stevens	-2.4

Table 4 Mean Martin-Quinn scores for the 2005-2007 terms. Note, negative scores indicate a liberal ideology and positive scores indicate a conservative ideology. The higher (lower) the number the more conservative (liberal) the ideal point is.

4.1 Methods

Corpus description

While the source and format of documents for this corpus is the same as that in Chapter 3, we selected a different timeframe. For this work, transcripts corresponding to cases from the February 2006 argument session (2005 Term) through the April 2008 argument session (2007 Term) were collected. This selection of cases represents a "natural court", a period of time during which the same 9 justices were in office with no changes in court membership. These justices include Chief Justice Roberts, Justice Stevens, Justice Souter, Justice Ginsburg, Justice Kennedy, Justice Thomas, Justice Alito, Justice Scalia and Justice Breyer. By using a natural court, we avoid potentially erroneous factors introduced by changes in court membership. Additionally, it increases our chances of avoiding the case where significantly less data is available for an individual justice due to factors external to that justice's behavior. While it would have been preferable to use more data, there is no longer natural court after the 2004 term; before then individual justices were not uniquely identified in argument transcripts. Of

the 179 cases argued during this period, 11 were held out due to inconsistencies in the database used for labeling each case.[16]

Case Segmentation

Cases were segmented into sequences of speaker labels. Each sequence was then divided into "speaker trigrams". Those familiar with the traditional view of trigrams, will recognize our interpretation of speaker trigrams. A speaker trigram is $S_iS_{i+1}S_{i+2}$ where S_i is the speaker of the i^{th} turn in the sequence (Manning and Schütze 1999). Figure 7 contains some example turns from the corpus (truncated for brevity), along with the sequence extracted from these turns and the resulting trigrams. We then obtained the count for each trigram across all cases and for all cases in each one of several conditions from the Spaeth database (e.g. direction of case decision, direction of Alito's votes, vote split, etc.).

[16] Held out cases include: 04-607, 05-204, 05-259, 06-1265, 06-1666, 06-618, 06-7517, 07-290, 07-330, 07-77 and 06-134 (*New Jersey v. Delaware*)

From *Snyder v. Louisiana* (06-10119)
CHIEF JUSTICE ROBERTS: Even though -- even though you're theory…
MR. BRIGHT: Oh, no.
CHIEF JUSTICE ROBERTS: -- that this jury did not return a…
MR. BRIGHT: No. Let me -- let me make this quite…
CHIEF JUSTICE ROBERTS: Thank you, Mr. Bright. Mr. Boudreaux…
ORAL ARGUMENT OF TERRY M. BOUDREAUX ON BEHALF OF THE RESPONDENT
MR. BOUDREAUX: Mr. Chief Justice, and may it please…
JUSTICE SCALIA: As to life imprisonment or as to the…
MR. BOUDREAUX: As to life imprisonment, Your Honor…
JUSTICE SCALIA: Where is this? I -- 364? Show me --
MR. BOUDREAUX: Beginning at 364 of the joint appendix…

Extracted Sequence:
ROBE L ROBE L ROBE START-ORAL L SCAL L SCAL L

Trigrams:
ROBE L ROBE, L ROBE L, ROBE L ROBE, L ROBE START-ORAL, ROBE START-ORAL L, START-ORAL L SCAL, L SCAL L, SCAL L SCAL, L SCAL L

Figure 7 Sequence of truncated turns, the sequence extracted from these turns and the resulting trigrams.

Labeling description

Labels were created using the Spaeth database. We experimented with variables along several dimensions including the direction of individual justices' and the Court's decision in cases (liberal/conservative) and the Court's vote split (5-4, 9-0, 8-1, etc). While we discuss only a sampling of charts in this chapter, all charts with greater than 10 cases for each variable value are included in Appendix A. In the Sections that follow we will cover the Vote Split (VOTE) variable, which contains the distribution of votes for a case, the Direction (DIR) variable, which contains the ideological direction of the case outcome and Justice Direction variables (*J*DIR) which contain the ideological direction of each justice's vote in a particular case.

The Rose Charts

Though radial plots have been explored extensively, use of radial plots for the visualization of sequential patterns and associated variables is a novel application of this layout (Draper et al. 2009). The outer ring of our diagrams (the petals in our terminology) is related to polar plots discussed by Draper et al (2009), while the inner ring is a pie chart. Because these charts are a novel application of radial layouts, we include the following technical description. For an explanation of how to interpret the charts, proceed to the Results Section (Section 4.2).

For each justice (except Thomas, again because of his infrequency of speaking) we created charts for all trigrams ending with that justice (i.e. all trigrams represented in a chart must end with the same S_{i+3}, where S_{i+3} is a justice). By concentrating only on those trigrams that end with the same justice, we can concentrate on turns that can be associated with "choice" on the part of that justice (i.e. the choice of that justice to speak after the speakers in the first and second positions in the trigram). We interpret this "choice" as the choice to *interact* with or pay *attention* to previous speakers. However, this is not necessarily the case; for example, these turns may arise if the justice is attempting to change the topic, and thus not paying attention to the previous speakers in the usual sense. Secondly, we chose to concentrate only on "typical" trigrams; because the vast majority of trigrams are of the form *JUSTICE LAYWER JUSTICE* or *LAWYER JUSTICE LAWYER,* all trigrams that did not have a lawyer in the second position were filtered out.

The center of each chart contains a pie graph representing the proportion of times the justice in the third position also spoke in the first position (i.e. $S_i = S_{i+3}$; "held the

floor" after the lawyer's turn).[17] Each of the outer petals represents one of the other justices that spoke in the first position (i.e. all other S_i). The width of each outer petal represents the frequency of each turn sequence normalized by the number of times S_i spoke, relative to the other petals. Thus, if the justice in the center devotes equal attention to all other justices (e.g. that justice follows-up on the same proportion of the turns produced by each other justice) all petals will have equal width. Because this looks at the proportion of turns rather than the count, the petals would be of equal width even if the frequencies of the sequences they represent are different. Petal radius represents the proportion of time with which two justices voted together, where shorter petals indicate the justices have more similar voting records than justices with longer petals. The inner dotted ring indicates 100% matching votes, and the outer edge of the chart area indicates 100% mismatch. Each object in the chart (petals and the pie graph) are colored on a gradient according to the proportion of cases in which that justices voted liberally or conservatively in the given category (i.e. that justices exhibited ideology), where white (blue in color versions) is liberal and gray (red in color versions) is conservative. We use counts of votes rather than Martin-Quinn scores because of the high variability of conditions chosen and because we want to represent the ideology within each condition. Note that because the range varies from condition to condition and because the range can often be quite narrow, the gradient is calculated within a condition, thus, a justice's color may vary from condition to condition. Finally, each petal is annotated with two values. The percent on the top, which is also in bold, is the width of the petal, while the percent

[17] We take the idea of "holding-the-floor" beyond the typical interpretation of maintaining control of a turn, to all instances where a speaker continues to produce turns after a single interceding turn from another speaker.

on the bottom represents the proportion of times that n-gram occurred compared to all other petals.

By representing turn-taking information in this way we hope to be able to capture broad patterns of the justices' turn-taking behavior. If we compare charts for different values within a condition, patterns may emerge that indicate a relationship between the values of that condition and a justice's behavior. For example, if we compare the turn-taking behavior of a justice when his or her vote is liberal to when the vote is conservative, and we note that a petal for a particular justice is short and narrow for liberal votes but long and wide for conservative votes, this could indicate that the justice in question has a greater tendency to follow-up on the particular justice of that petal in conservative cases. Furthermore, when the petal is long and wide, we may hypothesize that many of those follow-ups in some way challenge the justice of the petal since the length of the petal indicates the level of disagreement in the cases' outcomes.

4.2 Results

How to read the charts

Some of the patterns we discuss will be relevant either to wings of the Court or to justices from those wings. In these cases we will treat Kennedy, the swing justice, as irrelevant to these patterns. Additionally, we will identify speculative explanations for these patterns with *italic text* at the end of an observation.

Take, for example, Figure 8 "Stevens – Rose Diagram of All Cases". This chart contains all cases from our dataset. Because this chart is for Stevens, we find a pie chart in the center labeled Stevens, which indicates Stevens tends to "hold the floor", i.e. speaks again after an initial turn directed at the lawyer, ~75% of the time (signified by

the area filled in for the pie chart). It also shows that his voting record is one of the most liberal for this set of cases at, ~ 31% conservative votes (indicated by the color gradient).

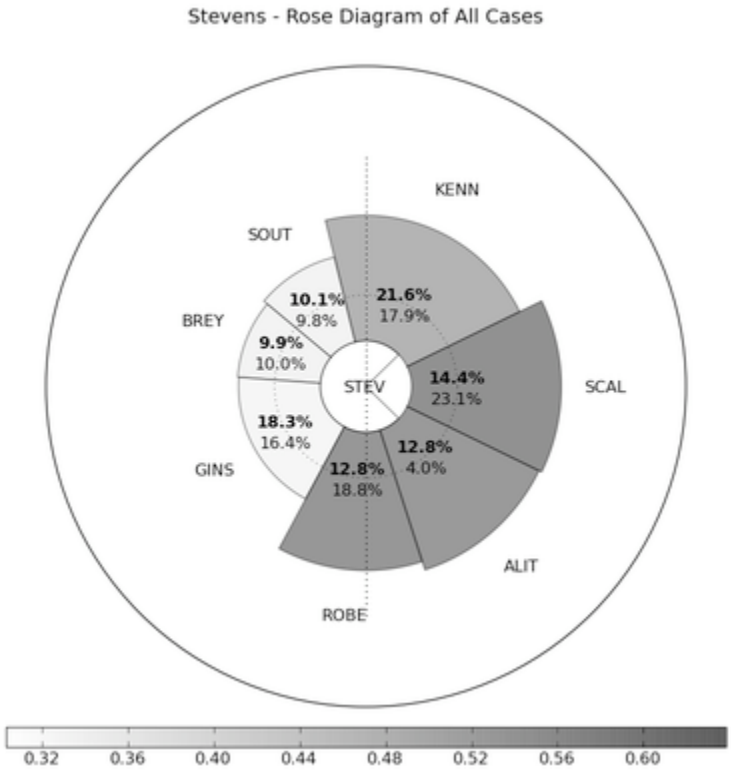


Figure 8 Stevens - Rose Diagram of All Cases

As discussed above, the outer petals represent all turn sequences in the dataset of the pattern $JUSTICE_1\ LAWYER\ JUSTICE_2$ ($J_1\ L\ J_2$) where $J_1 \neq J_2$, and in this case J_2 is Stevens. Thus, the petal labeled Kennedy represents all turn sequences of the form *Kennedy Lawyer Stevens*. The labels for this petal indicate that Stevens follows Kennedy 17.9% of the time when Stevens is not "holding the floor" and that the normalized proportion of this sequence is 21.6%. For Scalia, the relationship between these values is reversed, with the normalized proportion much lower than the unnormalized proportion. This indicates that while Stevens follows up on Scalia more often than he does Kennedy,

he does so on a smaller proportion of the turns produced by Scalia as compared to Kennedy. Finally, comparing the length of the Kennedy petal to the others, we see that Stevens votes with Kennedy less often than the liberal justices but more often than the conservative justices.

Looking at the outer petals we can make a number of generalizations, several of which are covered here in a top down fashion:

- Stevens has a greater tendency to follow-up on Kennedy, Scalia, Alito and Roberts (the justices he least often votes with) as a group than he does Ginsburg, Breyer and Souter (the justices he most often votes with).
- Holding Kennedy out as the swing vote, Stevens's interaction is much more evenly split between the conservative and liberal wings of the Court, with only slightly more follow-ups on justices he agrees with less often than ones he does agree with (40% vs. 38.3%). Thus, this indicates a somewhat disproportionate amount of attention given to Kennedy. *This may indicate that Stevens more often treats Kennedy as a "potentially persuadable justice", spending more time trying to convince him than other justices.*
- While the normalized proportion is fairly evenly spread out between the conservative justices in this chart, for the liberal justices, attention is skewed towards Ginsburg (18.3% towards Ginsburg vs. 9.9% and 10.1% towards Breyer and Souter). *This may indicate regular cooperation between Stevens and Ginsburg.*
- Of all justices Stevens is most likely to follow-up on Kennedy, at 21.6%, followed by Ginsburg at 18.3%.

- Finally, Roberts and Scalia both have much higher absolute percents compared to the relative percents, indicating that Stevens is less likely to follow-up on one of their turns despite a larger number of opportunities, indicating a greater proportion of turns go ignored from these justices.
- The absolute percent is much lower than the scaled relative percent for Alito, indicating a stronger tendency for Stevens to follow-up on Alito given the opportunity as compared to other justices, indicating Alito's turns are less often ignored as compared to Roberts and Scalia. *These last two observations together may indicate a tendency to argue with Alito more often than other justices in the conservative wing.*

Vote Split Condition (VOTE)

The VOTE variable in the Spaeth database indicates the distribution of the justices' votes (e.g. 5-4, 8-1, 9-0, etc.). Using this variable, we can test our intuitions about the sorts of patterns the charts will exhibit because we have well defined expectations for several features of the graph in this condition.

Figure 9, Kennedy – Rose Diagrams for 5-4 and 9-0 split cases, exhibits several patterns we would expect:

- 9-0 cases have maximal agreement between the justices; logically, if their decisions were unanimous then their votes always match.
- In 9-0 cases, justices always exhibit the same ideology. Their votes always match, thus their decisions have the same ideological direction.
- In 5-4 cases, Kennedy shows relatively high levels of disagreement with all justices, but slightly more agreement with conservative justices than with liberal

justices. We expect this pattern given that Kennedy is a slightly conservative swing justice, often casting the deciding vote in narrowly decided cases.

- In 5-4 cases, Kennedy exhibits an ideology in the center of the gradient while the other justices exhibit ideologies along the extremes of the gradient. This is what we would expect if Kennedy is the median justice and the other justices typically vote along their ideology in narrowly decided cases.
- Finally, in 5-4 cases, the petal width for Alito is very narrow, both compared to the other justices in 5-4 cases and compared to Altio's petal in 9-0 cases. Also, Alito has the shortest petal in 5-4 cases. *This may indicate that Kennedy tends to avoid interaction with the justice whose viewpoint is closest to his in narrowly decided cases.*

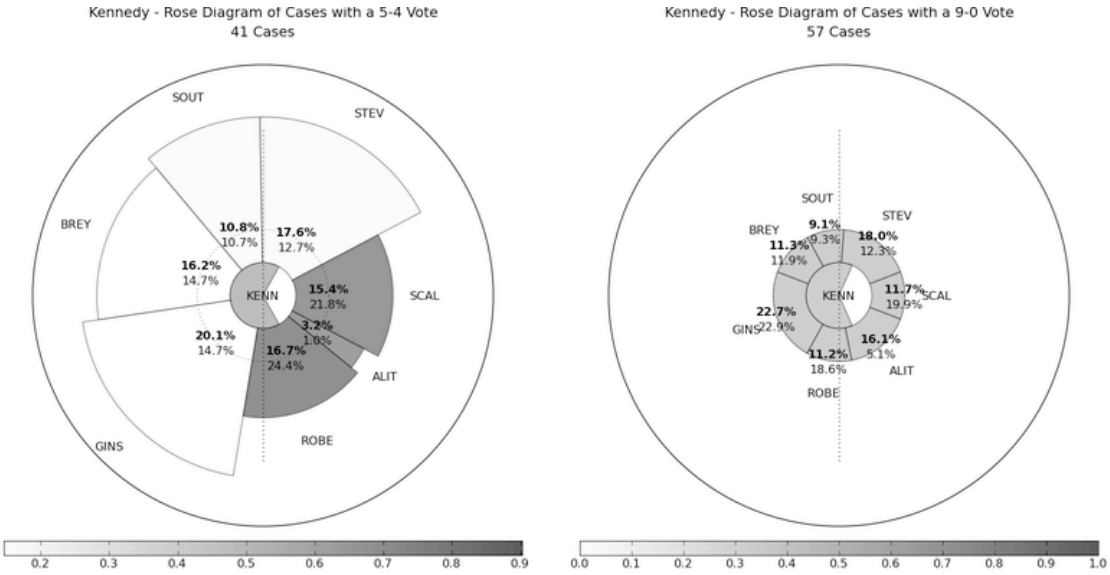


Figure 9 Kennedy – Rose Diagrams for 5-4 and 9-0 split cases

This pair of diagrams confirms our intuitions about the agreement and ideology patterns we expect to see when they are logically predictable. Additionally, the last bullet

point demonstrates the sorts of patterns that we can find when comparing levels of interaction across values in a condition.

Direction Condition (DIR)

The DIR variable in the Spaeth database indicates the ideological direction of a case's outcome. The ideological direction of a decision is determined based on the parties involved in the case and the issue area of the case according to the rules outlined in the Spaeth database documentation. Ideological direction is either liberal or conservative except in rare circumstances when no appropriate ideological direction can be determined. Below we discuss three diagram pairs in the DIR condition. In all charts, conservative decisions are on the right and liberal decisions are on the left.

Several observations can be made in Figure 10, Alito - Rose Diagrams for the DIR Condition (Alito is a conservative justice):

- When the eventual outcome of the case is conservative, Alito follows up on the liberal wing more frequently than when the outcome is liberal. This suggests a greater level of interaction via the lawyer between Alito and the liberal wing of the Court in cases that are eventually decided conservatively.
- There is less interaction between Alito and the conservative justices when the outcome is liberal as opposed to conservative. It should be noted that this is not the logical converse of the previous observation as the presence of a swing justices allows for changes in only one wing across a condition. *These two observations may indicate a slight tendency to argue more with justices that Alito disagrees with in cases where the outcome is likely to be against Alito's ideology.*

- These charts indicate an increase in interaction with Kennedy when the eventual outcome of the case is liberal. *For example, it is reasonable to assume that in any given case, each justice (in this instance, Alito) will have a fairly accurate expectation regarding the eventual outcome of the case. So, if Alito suspects that the eventual outcome of the case will be liberal (and especially if the case is likely to be split), Alito is likely to seek the support of Kennedy as a swing vote, which may likely be indicated as a higher degree of interaction.*

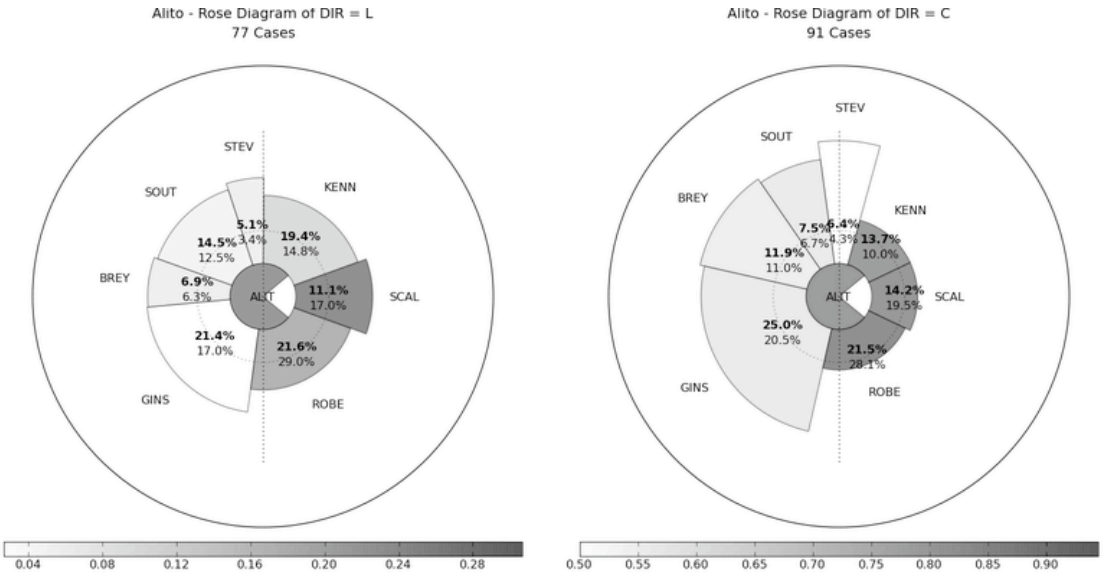


Figure 10 Alito - Rose Diagrams for the DIR Condition.

In the DIR condition for Ginsburg (Figure 11), we note the opposite basic patterns to those of Alito (Ginsburg is a liberal justice):

- In conservative cases we see a higher level of interaction with the liberal wing and a lower level of interaction when compared to liberal cases.
- We also see more interaction with Kennedy in conservative cases than liberal cases.

However, since Ginsburg and Alito are from opposing wings for the Court, these patterns can be used to form a single generalization. Namely, when the eventual outcome of a case is in opposition to the justice's general ideology, there is increased interaction with that justice's own wing, and decreased interaction with the opposing wing, as compared to cases when the outcome is inline with the justice's ideology. This pattern is observed for 5 of the 7 applicable justices (Kennedy excluded for the reason above and Thomas because he rarely speaks). Similarly, when a case's eventual outcome is against a justice's ideology, more interaction with the swing justice is observed than when the eventual outcome of the case is inline with the justice's ideology.

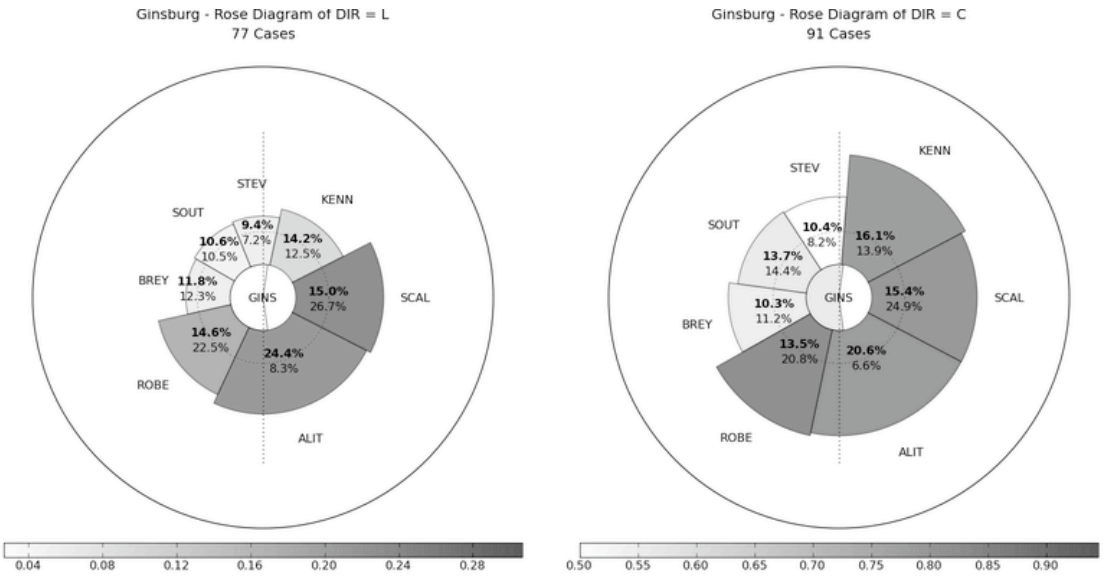


Figure 11 Ginsburg - Rose Diagrams for the DIR Condition.

In the above cases, Kennedy was treated as irrelevant to the patterns under discussion because he is the swing justice. Despite this, we can still make observations regarding Kennedy's interaction with the other justices. Figure 12 contains the DIR condition charts for Kennedy.

- Kennedy is more consistent than the previous justices we have discussed, when looking at his interaction with wings of the Court. He has only slightly higher interaction with the liberal justices in liberal cases and conservative justices in conservative cases. *We might expect this from a swing justice.*
- For each value in the DIR condition, for Kennedy there is a decrease in the proportion of follow-ups to the most liberal justice in that condition. That is, Stevens is the most liberal justice in cases with a conservative outcome while Ginsburg is the most liberal justice when the outcome is liberal; we see that Kennedy interacts with Stevens less when the outcome is conservative (i.e. he is the most liberal justice in conservative cases) and less interaction with Ginsburg when the outcome is liberal (i.e. she is the most liberal justice in liberal cases). *This could indicate a reluctance to get involved with the most extreme (liberal) viewpoint during a case.*

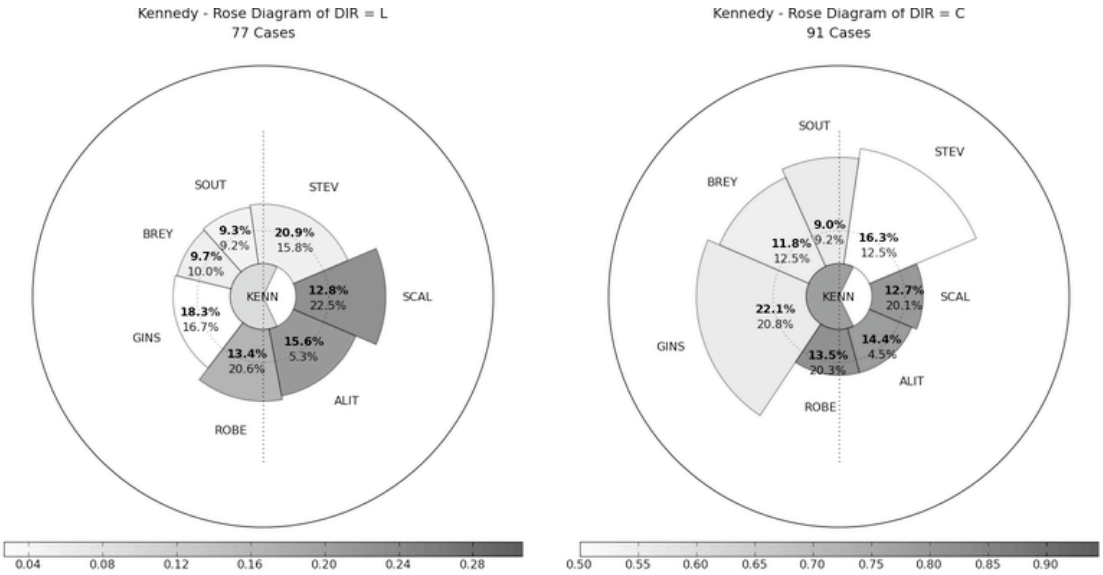


Figure 12 Kennedy - Rose Diagrams for the DIR Condition.

Justice Direction (JDIR)

Similar to the DIR condition, the *J*DIR condition has two primary values, liberal (L) and conservative (C); however, unlike DIR there is one *J*DIR value for each justice. So, ALTODIR (Alito's Direction) identifies the ideological direction of Alito's vote in a particular case. Note that no variable named *J*DIR appears in the Spaeth database, which instead contains one variable for each justice. We are simply using the name *J*DIR as shorthand for these variables. While other comparisons are possible, below we concentrate on charts comparing justices within their own *J*DIR condition. That is, for Alito we only present ALTODIR, for Breyer we only present BRYDIR, etc.

Figure 13 presents the two values for Alito in the ALTODIR condition. Note that because this is the ALTODIR condition, we expect that Alito will be on the extreme end of the ideology gradient in this case group (logically, if the value is conservative in the ALTODIR condition, 100% of the votes from Alito for that value will be conservative). We note several features in Figure 13 that may be interesting:

- First, when Alito's vote is liberal, there is a high level of agreement amongst the justices signified by the relatively tight radius of the outer petals. This indicates that Alito typically votes liberally only when most of the Court does so.
- When Alito's vote is liberal, we see a decrease in turns following the conservative justices and a slight increase in vote disagreement between these justices as compared to when Alito's vote is conservative. *This may indicate Alito has a tendency to follow-up more often with people who he agrees with.*
- For individual justices, we see some differences in the liberal wing. Though there is little change for Ginsburg and Stevens, we see notable changes in the relative

frequency when following Breyer (a decrease from the conservative to liberal) and Souter (an increase from conservative to liberal).

- We also note that the relative frequency of follow-ups on Kennedy shows a considerable increase from conservative to liberal. *Since Alito's record is more moderate than the rest of the conservative wing, this could suggest that Alito has more to discuss with the swing justice in particular when their interpretation of a case most closely aligned.*

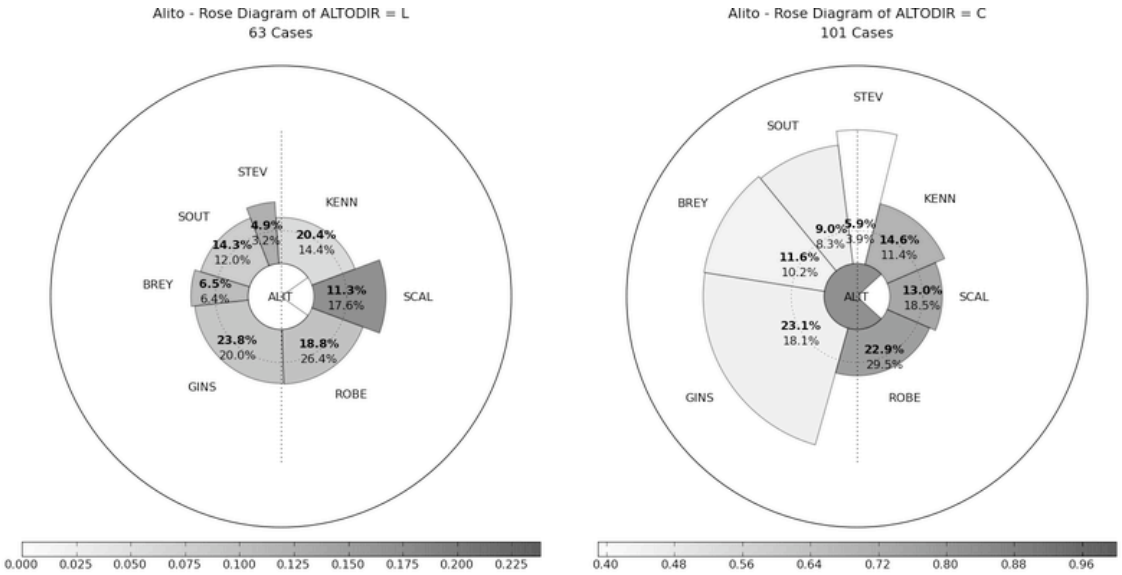


Figure 13 Alito - Rose Diagrams for the ALTODIR Condition.

Figure 14 contains the charts for Souter in the SOUTDIR condition. As in the DIR condition, it will be helpful here to look at things in terms of whether or not the vote matches the center justice's usual ideological direction, and whether other justices are from the same wing or the opposing wing (Souter is a liberal justice).

- Compared to Alito voting against his usual direction, we see a higher level of disagreement when Souter is voting against his direction. This indicates Souter's

conservative votes may be less closely related with conservative outcomes from the Court.

- As before, we see a slight increase in the normalized proportion of turns following justices from the same wing as the justice in the center when the case is against his typical direction (i.e. conservative).
- We also see a slight increase in the number of turns directed at the opposing wing when the outcome is against his usual direction.
- There is a decrease from conservative to liberal for turns following Ginsburg but an increase for turns following Stevens. We also see a fairly large decrease from conservative to liberal for Roberts and fairly small increases for Alito and Scalia. *These variations for individual justices likely suggest much more complex relationships between these justices.*
- Finally, we also see a relatively small increase from C to L for Kennedy, indicating relatively even amounts of attention given to Kennedy for both outcomes. *Perhaps this indicates that Souter doesn't use increased attention as a means of convincing another justice.*

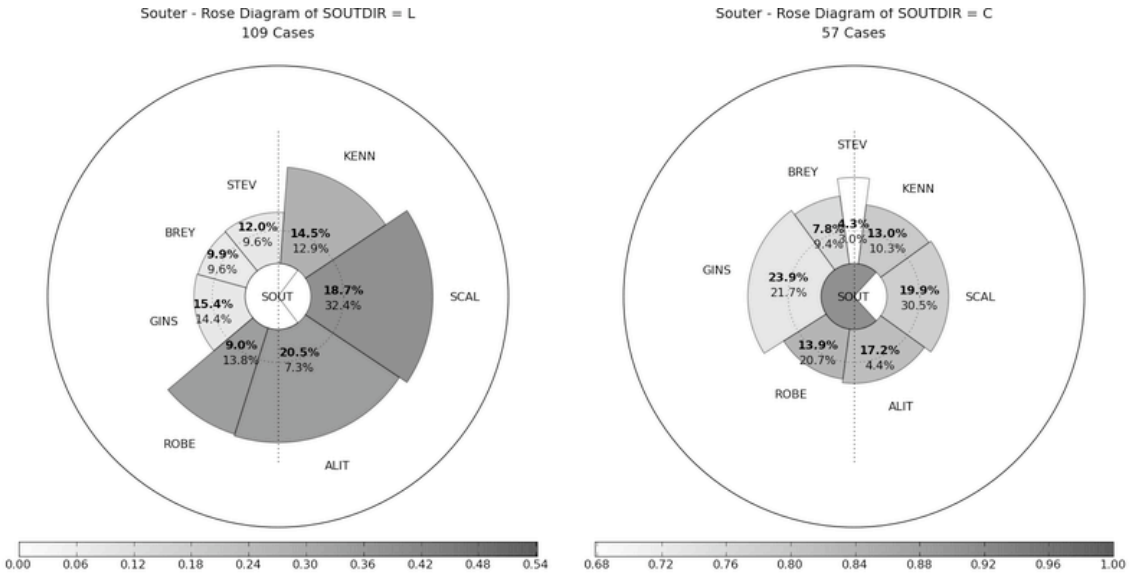


Figure 14 Souter - Rose Diagrams for the SOUTDIR Condition.

Unlike the two examples above, Kennedy's chart is fairly consistent with respect to the normalized proportions for each wing; however, we do still see small but potentially interesting differences between the two charts.

- When Kennedy's eventual vote is liberal, there is a slightly higher relative frequency of turns following liberal justices as compared to when his vote is conservative (the converse being true for conservative justices). *This suggests that Kennedy devotes slightly more attention to whichever wing he is likely to agree with.*
- It is also worth noting that for the conservative justices this difference primarily comes from a difference in the relative frequency of turns following Roberts, while for the liberal justices the difference is primarily distributed across Ginsburg, Breyer and Souter, with Stevens showing only a minimal change.

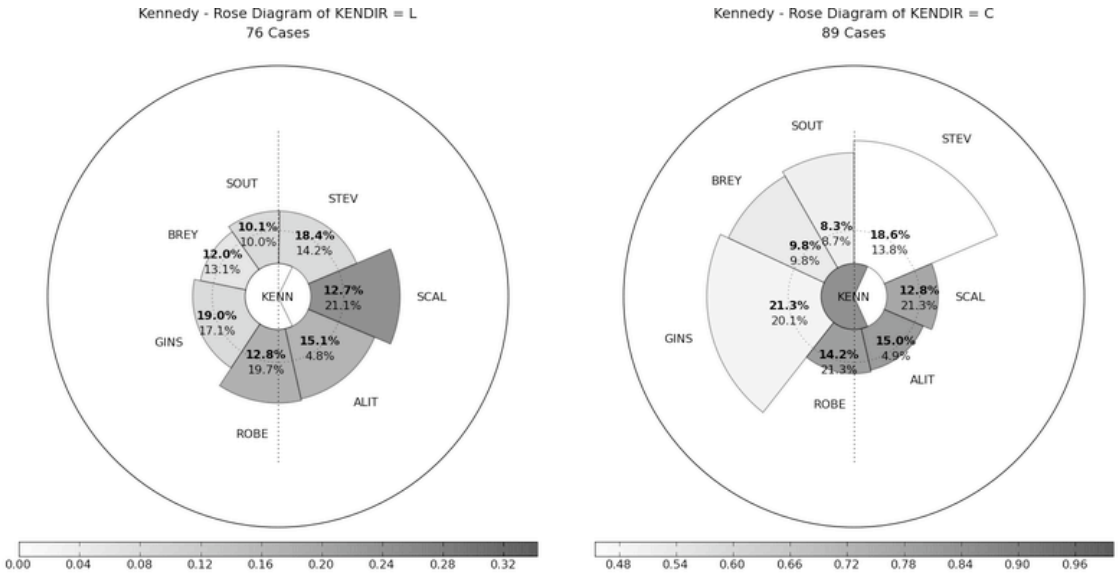


Figure 15 Kennedy - Rose Diagrams for the KENDIR Condition.

4.3 Discussion

The charts and observations above are a sampling of the sorts of general conversational patterns that can be observed for individual justices and the Court given outcome conditions that are of interest to legal scholars. For example, we saw a tendency of some justices from both wings to exhibit similar patterns to their respective opposing wings both in DIR and *J*DIR conditions. This suggests that there are patterns of turn-taking that can be associated with case outcomes, positively addressing the second point of this thesis.

Though we have only offered speculative explanations for these patterns, legal scholars should find that this sort of analysis could aid in the confirmation or discovery of patterns in the interactions of Supreme Court justices. Here we concentrated only on a particular subset of justices, outcome variables, and turn-taking patterns. While the appendix contains all justices for the conditions discussed above and several more outcome variables, there is no reason that these charts need to be limited to these

conditions. For example, it may be interesting to compare cases where a justice wrote a dissenting opinion compared to cases in which that justice did not. Or, one may wish to look at how patterns vary for certain case variables such as the lower court's direction, or combinations of variables such as unanimous conservative decisions.

The rose diagrams are also a novel application of radial layouts that can be used as a new tool for legal researchers when exploring the behavior of the Supreme Court. This approach is not limited to this particular pattern (i.e. $J_1\ L\ J_2$, where J_2 is held constant in the chart) of interaction either. There are numerous avenues for future research. For example, *L* could be broken down into petitioner and respondent or conservative party and liberal party.[18] If we are not particularly concerned with "choice" we may want to look at patterns that share a common J_1 or simply patterns that share a common justice in any position. The primary limiting factor in this sort of analysis is ensuring that one has enough cases for a good sampling of patterns. This was the primary reason we used a pattern that includes an additional individual between the two justices. Shorter patterns that include two justices are fairly rare, and longer patterns are sparser. However, with a careful selection of cases and relaxation of conditions one may still find that some patterns of this form can be examined as well.

[18] Where "conservative party" would indicate that a decision in favor of this party is a conservative decision, and vice versa for "liberal party".

Chapter 5 Vote Prediction

This Section describes our final set of experiments which build upon the insights revealed by the rose diagrams in the previous Chapter, examining vote prediction using turn sequences. If we can use turn-taking to forecast case outcomes, we will have demonstrated the validity of the third main point of this thesis; that the association between turn-taking patterns and case outcomes is predictive. Before discussing the approach, experiments and results, we will first briefly discuss our findings regarding the "most questions asked" method discussed in Chapter 2.

5.1 Prior approaches

We will first discuss our attempts to replicate results for the "most questions asked" rule discussed by Roberts, Shullman and Wrightsman (Wrightsman 2008), as well as Johnson et al. (2009a). While these projects leave the term "question" undefined, two reasonable interpretations exist. We could take question literally as any interrogative statement, which in the transcripts are usually identified with a question mark at the end. This sidesteps some of the issues discussed in Wrightsman, as transcription typically includes only one question mark per complete question, with no markings at the end of interrupted questions. However, we can also broadly define "question" as all statements produced by a justice. Though not the typical interpretation of what a question is, this seems to meet the typical treatment of turns produced by justices both as indicated in transcripts prior to 2004, which label the majority of Justice turns as "QUESTION", as

well as Wrightsman's example statements and Johnson's discussion of "attention given to a side". We explore both here.

Lacking the training data and some of the features used by Johnson et al. (2009a) we will use a simple rule based approach. We simply identify all *questions* in a case, take separate counts for each side and assign a "win" label to whichever side was asked the most questions. Following the lead of Johnston et al. (2009a), we can also apply both approaches to difference in questions asked and to difference in words directed at each side. By using a word based approach we again reduce the concerns about the definition of a "question". However, this does introduce other issues, such as the definition of a word (e.g. compounds, counting speech errors, contractions, etc.). To simplify matters, we take a word as anything separated by white spaces and word external punctuation (where characters such as apostrophe (') and hyphen (-) are word internal punctuation). Table 5 summarizes the results from these experiments.

Approach	Accuracy
Most Questions Asked (by turn)	56.8%
Most Questions Asked (by ?s)	56.8%
Most Words Used (by turn)	51.5%
Most Words Used (by ?s)	53.8%

Table 5 Comparison of "most attention given" approaches with varying interpretation of "question". "By turn" indicates that we count each turn as a "question". "By ?s" indicates we counted ?s in the transcribed justices' speech, usually indicating an interrogative statement.

As is clear, with this particular set of cases, no benefit is gained from a "most attention given" approach. As with most time periods, the majority of cases were reversed in this time period, creating a 65.6% most frequent outcome baseline which these approaches fail to meet. While interpreting "questions" as interrogatives

outperforms a turn based interpretation of questions on a "most words used" approach, no difference was found for the "most questions asked" approach. Moreover, the "most questions asked" approaches outperformed both "most words used" approaches.

Still, one could argue that the continued discrepancy over the power of a "most questions asked" rule is a problem of sample size. In the case of the smaller manual studies, high accuracy may simply be attributed to a favorable sample selection. For the larger study, the distribution of questions compared to case outcome provided by Johnston et al. (2009a) is unambiguous, and clearly demonstrates that at least in the extreme cases this rule does appear to be valid. Models trained on a larger sample will have a more representative distribution of these extreme cases. In fact, like Johnson et al. (2009a), if we assign labels based on the "most attention given" rule for extreme cases and use the majority class for the rest we do get similar accuracy. Results provided in Table 6 are for cases in which the difference in number of questions or words addressed to a side is more than 2 standard deviations from the mean.

Approach	**Cases**	**Accuracy**
Most Questions Asked (by turn)	8	87.5%
Most Questions Asked (by ?s)	7	75.0%
Most Words Used (by turn)	6	83.3%
Most Words Used (by ?s)	6	60.0%

Table 6 Comparison of "most attention given" rule for extreme cases (i.e. difference in words or questions is > 2 s.d. from the mean). The "Cases" column indicates how many cases met this criterion.

Because "extreme cases" are simply those that have differences in attention (measured by word or turn counts) given to a side more than two standard deviations. It may be possible to identify these cases in advance by examining the distribution of prior

cases and determining whether or not the difference in attention given for each new case is within or outside two standard deviations for the distribution of previous cases.

5.2 Forecasting votes

In our discussion of forecasting oral argument transcripts attention must be given to both the sorts of features used and the outcomes that we are forecasting. We focus on using features that are easily extracted automatically, with little to no human input. Instead of concentrating on the content of the oral arguments, we concentrate on the conversational dynamics of the justices and lawyers involved in a case, as a function of their turn-taking behavior. While the content of justices' and lawyers' turns is very likely informative about a case's outcome, several factors make it difficult to utilize content with automatic methods. First, because the transcripts are composed mostly of spontaneous conversation, performance of existing natural language processing techniques such as parsing and even POS tagging is considerably lower than in tasks where the input is written text or even prepared speeches. Second, while features explored in some manual forecasting approaches such as "hostility" and "sympathy" are certainly present in the content, these features are also not well defined and not easily identified using computational methods. Those features that are somewhat more easily identified, such as topic area, vary widely from case to case. This makes it difficult to find a relationship between these easily identified features and the cases outcome. Finally, as we have shown above, because simple turn based "most questions asked" or "most words used" are limited to extreme cases, their recall (in this instance the proportion of correct predictions to the number of cases) will be low despite high

precision (in this instance, the proportion of correct predictions to the total number of predictions).

One important consideration when predicting case outcomes is deciding just what outcome one wants to predict. The most obvious choice, and the one most often chosen in previous prediction tasks, is whether a case will be affirmed or reversed. There are, however, other potentially relevant options to choose from. For example, justices are very rarely spoken of in terms of their tendency to affirm cases. Typically, when examining justice's voting records, one wants to speak of justices in terms of the direction of their ideology; either liberal or conservative. While the vast majority of cases are either affirmed or reversed, typically each of these decisions is liberal or conservative as well. If the most relevant dimension for discussing justices is the direction of their ideology, then it seems fair to at least consider prediction of case outcomes along this dimension as well. For these reasons, conservative vs. liberal was the primary outcome feature we concentrated on.

However, as one would expect, conservative and liberal outcomes do not occur with equal probability, and so the baseline for such a condition is not 50%. However, we can achieve a 50% baseline by splitting cases and then viewing outcomes as a win or lose variable for each side of the case. We explore this outcome in our third experiment.

5.3 Methods

Corpus Description

We use the same corpus as used for the rose charts, described in Section 3.1.

Turn Distribution

As with the sequence prediction task in Chapter 3, from each case we extracted speaker IDs and meta-symbols from the transcript. As before, litigants were reduced to a single symbol (reported here as *L*). To conserve space when reporting tables, justices are identified by the first four letters of the justice's last name (Table 7). From each sequence we then counted all turn 4-grams. Since the objective of this experiment is to leverage justice interaction as a means for predicting case outcomes, we don't want the n-grams to be too short. If the n-grams selected are too small we risk losing information about the interaction between justices (as the typical sequence of speakers is Justice, L, Justice, L,….). If the n-grams are too long, however, we begin to face sparseness problems, since the larger *n* gets the more variability there is and thus the lower the counts will be. Thus 4-grams seemed to be the ideal selection.

Speaker	Symbol	Count
Non-justice party	L	19840
Chief Justice Roberts	ROBE	3890
Justice Stevens	STEV	1964
Justice Scalia	SCAL	4277
Justice Kennedy	KENN	2196
Justice Souter	SOUT	2590
Justice Thomas	THOM	3
Justice Ginsburg	GINS	2379
Justice Breyer	BREY	2668
Justice Alito	ALIT	840

Table 7 Speakers and their corresponding symbols. The count column identifies the frequency with which each symbol appears in the corpus.

There are 41,417 occurrences of 1,072 unique n-grams. Table 8 summarizes the 20 most frequent 4-grams in the corpus. Because justices do not frequently speak in adjacent turns, after each justice's turn there is typically a lawyer's turn. Because of this, n-grams usually occur in corresponding pairs that have in common a *Justice Lawyer Justice* trigram, but differ in whether the four-gram starts or ends with a lawyer. We, therefore, report these pairs together. However, note that they do not always rank next to each other, and so the Table is ordered by the rank of the most frequent 4-gram in the pair.

Corresponding n-grams	Counts	Ranks
L **SCAL L SCAL** / **SCAL L SCAL** L	2467 / 2456	1 / 2
L **ROBE L ROBE** / **ROBE L ROBE** L	1801 / 1651	3 / 8
L **BREY L BREY** / **BREY L BREY** L	1746 / 1726	4 / 6
L **SOUT L SOUT** / **SOUT L SOUT** L	1729 / 1705	5 / 7
STEV L STEV L / L **STEV L STEV**	1237 / 1220	9 / 10
KENN L KENN L / L **KENN L KENN**	1182 / 1158	11 / 12
GINS L GINS L / L **GINS L GINS**	1137 / 1122	13 / 14
L **SCAL L ROBE** / **SCAL L ROBE** L	418 / 337	15 / 18
ALIT L ALIT L / L **ALIT L ALIT**	397 / 387	16 / 17
L **ROBE L SCAL** / **ROBE L SCAL** L	331 / 328	19 / 20

Table 8 20 most frequent n-grams grouped by correspondence pair, ranked by most frequent n-gram in pair

Note that the majority of these 4-grams include justices "holding-the-floor" with the only two instances of more than one justice in the bottom of the table. Despite the fact that the most common 4-grams follow this pattern, many less frequent n-grams represent three or four instances of a justice speaking (Table 9).

n-gram	count
BREY BREY L BREY	18
SCAL BREY L BREY	18
SCAL L SCAL SOUT	16
SCAL L SCAL SCAL	16
SOUT L SCAL GINS	15
BREY SCAL BREY SCAL	5
ROBE SCAL ROBE SCAL	3
KENN GINS ALIT GINS	1

Table 9 Infrequent *n*-grams containing 3-4 instances of justice turns.

Note, because the conversational patterns of the Supreme Court are usually very consistent, rare patterns like those in Table 9 often indicate uniquely transcribed events; the majority of instances where the same justice has two adjacent turns in the transcript indicate laughter in the Court. When two justices' turns are adjacent to one another this usually indicates an interruption has occurred. Figure 16 contains examples of both laughter and interruptions from the corpus. In the first excerpt, there is laughter after Breyer's first turn, after which he continues to speak.[19] Thus the sequence is transcribed as BREY BREY L BREY. Also note, Mr. Sorrell's turn ends with a "--" indicating that his turn was unfinished. We interpret this as an interruption. However, because Mr. Sorrell is the attorney in this instance, we do not observe anything unusual in the sequence for this pair. In the second excerpt the transcript indicates that Roberts was interrupted by Scalia, after which Roberts attempts to "hold-the-floor" by interrupting Scalia, but eventually gives way to a second interruption by Scalia. This sequence is then transcribed as ROBE SCAL ROBE SCAL.

[19] It is unclear from the transcripts whether this laughter should be attributed to Justice Breyer or someone else.

Randall v. Sorrell (04-1528)
JUSTICE BREYER: No, no. It's $200. Coffee and donuts are expensive. (Laughter.)
JUSTICE BREYER: Okay? Count it or not?
MR. SORRELL: We don't -- our coffee is not that expensive, but --
JUSTICE BREYER: Donuts and coffee. In other words, it counts as long as it's over $100.

Samson v. California (04-9728)
CHIEF JUSTICE ROBERTS: What about --
JUSTICE SCALIA: Is --
CHIEF JUSTICE ROBERTS: What about --
JUSTICE SCALIA: Is that right? I mean, even in prison, I -- what -- I'm not sure you could even do that if they were still in prison. Can you subject people in prison --

Figure 16 Examples of "Laughter" and interruptions in the transcript

Data Preparation

Before proceeding with any sort of classification, several preprocessing steps were taken in some experiments in order to address sparseness issues as well as remove irrelevant and potentially distracting features:

- All non-justice parties are reduced into a single symbol. Since these are most often attorneys, we reduced them to the L symbol. This step was taken for all experiments.
- Eliminate all turns not ending with a justice. This essentially reduced the presence of feature pairs of the type discussed above.
- Remove all n-grams containing markup, including TIME, as well as the special symbol for the beginning and end of a case.

- Collapse all justices into one of three categories; liberal (occupied by Stevens, Souter, Ginsburg, and Breyer), conservative (occupied by Roberts, Scalia, Thomas, and Alito) and swing (occupied by Kennedy).

While not taken in all experiments, as it seemingly disregards quite a bit of information, this final step deserves some more attention. The motivation behind such an approach is that it greatly reduces sparseness in the data. Not only is the liberal/conservative ideology one that is more or less common knowledge, often observed both in scholarly literature and in the media, but it is also clearly indicated in each justice's voting records. Moreover, ideology is often considered one of the more relevant dimensions over which a case is decided, so it is extremely relevant to predicting case outcomes. Even when the outcome to be predicted is affirm/reverse or agree/disagree, the interaction of the liberal justices and conservative justices with the swing justice can be informative in predicting case outcomes. However, rather than capturing the interaction between individual justices, this is more accurately described as capturing the interaction between wings of the Court. Given the rose charts, we may hypothesize that this interaction between the wings is also a relevant point to look, as patterns were observed in the way that members of each wing treated opposing wings. That is, patterns at the "wing level" should be relevant.

In addition to these data preparation options, we also calculated feature values in two ways. The first, and most straightforward, was to simply use the absolute counts of each n-gram. For the second approach we used relative feature scores. For each n-gram we divided its frequency by the count of all n-grams for that case. The denominator included all n-grams; i.e. even those that were removed from the feature set using the

filters described above. While the feature values do not sum to one this means we will be able to indirectly encode potentially useful information such as case length.

Baselines

In most studies predicting Supreme Court outcomes, little attention is given to baselines. Understandably, at first blush, when trying to predict an outcome like affirm or reverse a 50/50 baseline seems applicable. There are only two outcomes in general (others are possible, but rare) and both seem to occur with a fair amount of regularity. However, when examining the history of the Court, one finds strong tendencies for certain outcomes to occur more often than others. Needless to say, the Supreme Court is not as simple as a fair coin toss. So, we need to consider the frequency with which each outcome occurs in each condition in order to establish more reliable random baseline.

For an affirm/reverse condition we look back at the frequency with which the Court upheld the lower court's decision and the frequency with which the lower court was overturned. In doing so we find that the Court has a tendency to reverse cases more frequently than it affirms cases. Taking a sample of 1000 cases from the 1997 term to the 2007 term, the Court affirmed cases 34.4% of the time and reversed 65.6% of the time. Over shorter periods this tendency can shift drastically; for example, if we look at a 20 case "moving average" of affirm decision chronologically over this time period (based on date of argument) we see that the average reaches as high as 100% and as low as 35%. Thus, a random baseline for this example is not 50/50.

At first this may seem surprising; however, one must consider how cases are selected. Of the approximately 9000 cases submitted to the Court each year, only 80 or so are selected to be heard by the Court. Naturally, then, the justices are picking those cases

which they view as most important, and as it turns out there is a slight bias for those cases which the Court will overturn.

For a liberal/conservative baseline, the Court is a bit more balanced, at 54.2% conservative and 45.8% liberal for the Roberts court (with Alito). This likely has more to do with the composition of the Court than anything else. In fact, one might expect to see a court with a conservative chief justice and a slightly conservative leaning swing vote with a greater proportion of conservatively decided cases.

Despite these unbalanced baselines, it is possible to construct experiments that do have true 50/50 baselines. The experiment labeled *The Court II* is an example of this. By splitting the case into sides, (i.e. all turns during petitioner's argument is one side, all turns during respondent's argument is another) and setting the outcome to win/lose we ensure that there are an equal number of win instances in the data as there are lose instances (as for each case one side must win and the other must lose; again, except in rare circumstances).

5.4 Experiments

We discuss four experiments in this Section, three dealing with classification of the Court as a whole (*The Court I, The Court II and The Court III*) and one dealing with the classification of Thomas's votes (*Thomas*).

The Court I: The first experiment conducted in this category attempted to predict whether the Court's ruling would be liberal or conservative. We found that for this sort of task, predicting the outcome of a case for the Court, classification was highly sensitive to sparseness, so we collapsed justices into Liberal, Conservative and Swing categories. We also employed the filter that reduces the presence of pairs. We use absolute rather than

relative feature values. Classification was conducted using the LIBSVM 2.86 implementation of support vector machines (SVM) with default parameter settings 5-fold cross validation and parameter tuning (Cortes and Vapnik 1995).[20]

The Court II: As a second experiment we tested the "in favor of side" condition. While somewhat more artificial than other experiments, this approach does allow us to examine these features in a truly balanced context. We prepared the data by splitting each sequence by side, so each case was composed of two sequences; turns produced during the petitioner's arguments and turns produced by the respondent's arguments. Because the Court has a relatively high affirm baseline (meaning the Court usually votes in favor of the petitioner) we removed all information about the side that was being spoken to from the feature set which are introduced in the form of meta-symbols. By splitting the data, we also magnify the sparseness problems from before, and so we continue to collapse justices into their ideologies. However, also because of the high level of sparseness, we did not remove n-gram pairs, as doing so often reduced the features in any given case too far. This experiment used relative rather than absolute feature values. Again, note that since in each case one party must win while the other loses, this ensures that there are an equal number of winners and losers in the dataset. Again we used the LIBSVM implementation of SVMs with default parameter settings and 5-fold cross validation with parameter tuning.

Unlike the liberal/conservative classification, the choice to collapse justices into liberal, conservative and swing categories for this condition might at first seem like an irrelevant dimension on which to reduce sparseness. However, there are some important points to keep in mind. While the Court for this corpus was balanced with liberal and

[20] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

conservative justices (4 of each), as a result of Thomas's general silence, the number of speakers from each wing is unbalanced. Moreover, looking at the wings rather than individual justices, it may be the case that we are able to capture instances of the "three-way" conversation described by David Frederick's where the justices are conversing both with each other and with a particular lawyer (Biscupic 2006, Johnson et al. 2009a). To see why this may matter, consider the rose diagrams discussed in Chapter 4. Although we remove identity information of justices by collapsing the data, we are able to maintain the general effects that have to do with wings of the Court, and since Kennedy is the only swing justice, no identity information is lost for this justice. As a result, we may see cases where either Kennedy is showing high levels of agreement with a particular wing, or where the wings are jostling for support from Kennedy.[21] In either situation, this may be an important factor as the swing vote will often be the deciding factor in a case.

The Court III: In addition to SVM approaches, in these conditions we also attempted some rule-based classification conditions. This allows us to identify n-grams that are most informative in classification, thus giving us a way to search for those exchanges between justices that may be particularly helpful in identifying the outcome of a case. This experiment used the WEKA 3.6.0 J48 implementation of decision trees.[22] We found that our original data preparation options did not perform well with decision trees, however, after experimenting with other data preparation options we found that by only collapsing justices into their ideology some improvement over baseline was achieved.

[21] In order to test whether we were simply predicting Kennedy's votes in this situation we tested classification for his votes, for or against a particular side of a case, with the same settings. The classifier achieved 58.3% accuracy which suggests this was not the case.

[22] http://www.cs.waikato.ac.nz/ml/weka/

***Thomas*:** Thomas's voting history indicates a relatively high baseline at 69.5% conservative votes. This, of course, is unsurprising given that Thomas is often considered one of the most conservative justices currently on the Court. What is surprising is that despite this relatively high baseline and his tendency to almost never speak during oral arguments, we are able to use the approach described above in order to gain insight as to when Thomas will cast one of his relatively rare liberal votes. For the experiments with Thomas we found that by not reducing justice IDs to their liberal/conservative classifications and by using only those n-grams with more than one justice we did see a reasonable improvement in Thomas's classification accuracy. We used relative rather than absolute feature values. Classification was conducted using the WEKA 3.6.0 implementation of Decision Tables (Kohavi 1995).

Results

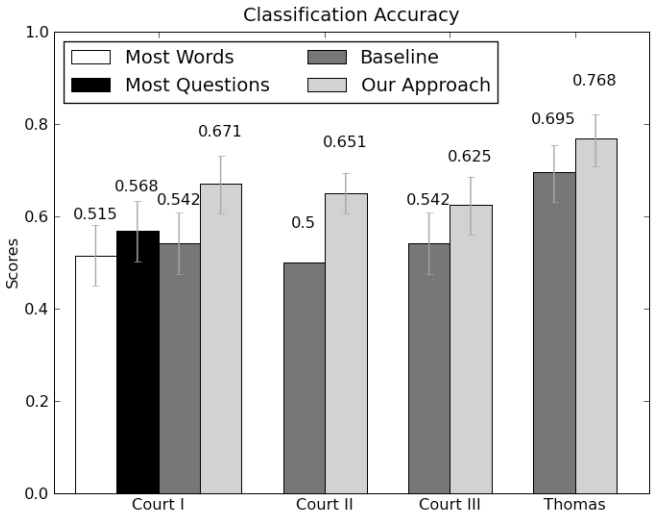


Figure 17 Classification results including prior approaches (Court I only), baseline, and absolute accuracy. Error bars are the 90% confidence interval as calculated by the Clopper-Pearson method for inferring exact binomial confidence intervals.

The results of the experiments are detailed in Figure 17. Error bars are calculated as the 90% confidence interval as computed by the Clopper-Pearson method for inferring exact binomial confidence intervals (Clopper & Pearson, 1934). We compare our results to prior approaches for *The Court I*, and the baselines described above for all experiments. In all cases, our approach outperforms both prior approaches and the baseline. However, as indicated by the error bars, confidence intervals overlap in several instances. Both *The Court I* and *The Court II* outperform the baseline at a 90% confidence level. We also see that *The Court I* outperforms the "most words used" approach on this dataset. This is an important finding because the "most words used" approach was found to be the most powerful approach in prior studies (Johnson et al. 2009a). Moreover, we see that these results are comparable to experiments that used an order of magnitude more data (Johnson et al. 2009a). For all experiments on the Court, we found that collapsing justices was a very useful preprocessing step. The greatest increase in accuracy was provided by SVMs, regardless of the condition. And of the two experiments that used SVMs the greatest increase was over the split-case baseline of 50%. While decision trees do not provide the double digit increases that SVMs do, they still provide some improvement over baseline with the added benefit of providing decision trees that can be examined. The results for Thomas are perhaps the most surprising. Though the improvement is relatively small, not only are we dealing with a much higher baseline, but this suggests that the interaction of the justices who do talk during cases is correlated with the way Thomas will vote even though he rarely participates in oral arguments.

Because the decision tables are easily interpretable, we can also examine the specific n-grams that are most informative in classification. We are especially interested in n-grams that contain more than one justice, because these best highlight the interactions between individual justices. Decision tables returned four such 4-grams that contained more than one justice. Figure 18 contains these sequences along with examples of these sequences from the corpus.

BREY BREY L GINS
Ex. From *Michael A. Watson v. United States* (06-571)
JUSTICE BREYER: I don't want to put you in a whipsaw here. (Laughter.)
JUSTICE BREYER: Sometimes policy seems relevant, too, to figure out what Congress wanted. But let me go back to the question I had, which is do you want to us overturn Smith?, Are you asking that, because I could understand it more easily if you said, look, both sides of the transaction should be treated alike, but they should be both outside the word "use."
MR. KOCH: I do not believe it's necessary for this Court to overrule Smith in order to rule for the Petitioner here, because of -- because of the differences, first of all linguistically; and secondly because of the reliance on Bailey.
JUSTICE GINSBURG: And in answer to my question, you said you were not urging the overruling of Smith?

SOUT SCAL L SCAL
Ex. From *Federal Election Comm'n v. Wisconsin Right to Life, Inc.* (06-969)
JUSTICE SOUTER: And it is impossible to know what the words mean without knowing the context in which they are spoken.
JUSTICE SCALIA: When the Government put these exhibits, were those exhibits complete with context?
MR. BOPP: No. There was no --
JUSTICE SCALIA: I didn't think so. They just -- they just -- what the ads were.

SCAL L SCAL GINS
Ex. From *Engquist v. Oregon Dept. of Agriculture* (07-474)
JUSTICE SCALIA: That's certainly an equal protection. She could be fired at will and everybody else can be fired at will.
MS. METCALF: Agreed.
JUSTICE SCALIA: Why isn't that equal protection of the law?
JUSTICE GINSBURG: Except this wasn't -- this wasn't employment at will, right?

BREY ROBE L GINS
Ex. From *Travelers Casualty & Surety Co. of America v. Pacific Gas & Elec. Co.* (05-1429)
JUSTICE BREYER: And, and yet there are no briefs from them; there are no -- there is no article that I could find in Bankruptcy Journal.
CHIEF JUSTICE ROBERTS: Well, there may be no briefs from them because it isn't the question on which we granted cert, is it?
MR. BRUNSTAD: Chief Justice Roberts, that's Official correct. And our view is that the Court should deal only with the Fobian rule. And the alternative argument which Respondent presents was never argued below, was not decided below, was not presented in the opposition to certiorari. It's been rejected by every single court of appeals --
JUSTICE GINSBURG: But it would be proper to remand for the Ninth Circuit to consider those other arguments?

Figure 18 Informative sequences from *Thomas* decision trees with examples from transcripts.

Since the baselines for individual justices are so high, any improvement in classification accuracy is going to come from the ability to predict unusual behavior from that justice. This is just what we found in the case of Thomas. One can already predict the majority of Thomas's votes simply by assuming his vote will be conservative. In order to move beyond this simple baseline, one needs to be able to predict liberal cases. By predicting these with high precision, we are able to boost performance when predicting outcomes for Thomas. Though such results may be subject to the danger of over-fitting, as additional cases are being created, it will be possible to test this approach further. Of course, as justices change so too will the performance this approach.

Discussion

These classification experiments built upon the observations in Chapter 4 that turn-sequences are associated with case outcomes. These results indicate that there are patterns in justices' turn-taking behavior that are in fact predictive of case outcomes. Additionally, we show improvement on our dataset over approaches previously shown to have the best performance the most comprehensive prior study. Moreover, the accuracy is comparable to that of studies that used an order of magnitude more data than our study, while exploring a novel hypothesis about the predictability of Supreme Court outcomes and the features of the case that are used make predictions (Johnson et al. 2009a).

The fact that any benefit at all is achieved using interaction features as simple as turn-taking is a novel finding that may surprise some researchers (Evans, M. personal correspondence, August 28, 2009). Questions still remain as to why the features used are important. Without a doubt the content of justices' turns are informative with regard to a case's outcome, but what about the conversational nature of the exchanges represented by

our features? Future research might ask what characteristics of these exchanges are informative. Perhaps it is general features, such as the tone of the exchange, or perhaps these n-grams isolate strategic exchanges where judges in opposition to one another are looking to counter other justices' arguments and judges in agreement to one another are providing support.

Interestingly, this approach has the potential to predict both the behavior of the Court as well as individual justices. This is an important finding as it suggests that these approaches may not need to be restricted to natural courts.

This work represents a methodologically novel approach, thus creating a new tool for researchers looking to gain a greater understanding of the Supreme Court and the justices. As discussed below, as more data is created (thus reducing sparseness) numerous extensions to this approach present themselves, suggesting the possibility of richer more powerful models of justice interaction and court behavior.

Chapter 6 Conclusions

This work represents the first steps towards modeling the relationship between Supreme Court justices' interactions and actions. We have novelly applied computational methods for pattern discovery in Supreme Court discourse which may more generally be applied in legal discourse. While legal scholars and other court followers may have intuitions about the social dynamics of the Court, these intuitions are most often limited to a few areas of expertise and a narrow range of examples. What this work offers is a global approach to pattern discovery in the social dynamics of the Supreme Court justices. With these patterns, legal scholars are given a new avenue for research that can lead to a greater understanding of this country's highest court that would otherwise go unexplored.

This work addressed three objectives: to show that a) predictable high level patterns exist in the conversational dynamics of the Supreme Court, b) these patterns may be associated with other areas of interest to legal scholars such as voting patterns of the justices, c) this association between linguistic patterns and judicial patterns may be utilized both to provide short term insights (i.e. predicting the outcome of a particular case) and deeper insights about the behavior of the Supreme Court. Our results indicate that a, b and c do hold. We have found that by combining features with regard to turn content, discourse marker use, and personal reference we can gain information about who is speaking when and that by increasing the history of these features we can further boost the reliability of these methods. The rose charts demonstrate that interesting patterns can

be observed when we are looking at summaries of the turn-taking behavior for various conditions. Our prediction approach performed significantly better than prior approaches on the same data and comparably to approaches utilizing an order of magnitude more data (Johnson et al. 2009a). These results indicate that turn-taking patterns are in fact predictive of case outcomes.

In addition to the contribution of positive results, we have also made a number of methodological contributions as well. While the analysis of Supreme Court discourse is not new, our approach of viewing the patterns of Supreme Court turn-taking as both predictable and predictive of case outcomes is a novel one, and we have offered several techniques to explore this hypothesis. We addressed only a narrow range of questions with these techniques, but expect that legal scholars will find a wide array of hypotheses to explore. Additionally, our rose diagrams are a new application of radial plots that are helpful in visualizing the relationship between turn-taking sequences and actions (Draper 2009).

6.1 Future work and Unanswered Questions

Unfortunately, sparseness is a major limiting factor in combining content with turn sequences for the Supreme Court. However, as data is continually being created, these problems should be continually reduced. Moreover, though not explicitly identified in the transcripts prior to 2004, the identity of individual justices is not lost, as the audio transcripts of these cases still exist. Perhaps by combining audio speaker recognition techniques with our justice identification approach, one could reconstruct speaker identities for these earlier cases (Yuan and Liberman 2008). Doing so would provide considerably more data for experimentation. If sparseness issues are appropriately

addressed one could incrementally increase the amount of information used in turn sequences. For example, with limited additional work, one could include further turn features such as interruptions, perceived humor (indicated in transcripts with a "laughter" marker), and question vs. statement. As indicated above in Section 5.3, while not overtly marked, these first two features still managed to find their way into our dataset as discussed above and were some of the most informative features in classifying Thomas. While overtly marking these features increases sparseness too far, adding more data reduces this problem making the overt marking of these features viable; and given the results above one would expect them to be helpful. As other researchers have found, the questioning pattern is likely indicative of case outcomes, at least in extreme cases. Thus, one might expect some benefit from incorporating questioning features in the turn sequence.

Moreover, in many cases the existence of interruptions and laughter is indicative of higher level features of a turn, such as hostility and tone of questioning. Though the reliability of identification of these features is currently untested, work in areas such as sentiment detection may be useful in attempting to identify these features (Pang and Lee 2008). If successful, these too could be included in the turn sequence and would likely give further insight into the interaction of the justices.

Another strong cue to the interaction of justices would be the discourse relations that hold between justices' turns. Again, while incorporating features for discourse relations in the turn sequence would inherently increase sparseness, if and when sparseness is addressed, including discourse markers in the turn sequence is a logical first step to creating a richer feature set that includes information about discourse relations.

Ultimately, one would ideally want to identify the underlying relations that hold between the turns in the sequence. Identifying the speaker or wing of the speaker along with how the turn relates to the previous turn would clearly provide rich information about the interaction of justices and would likely be highly informative regarding case outcomes.

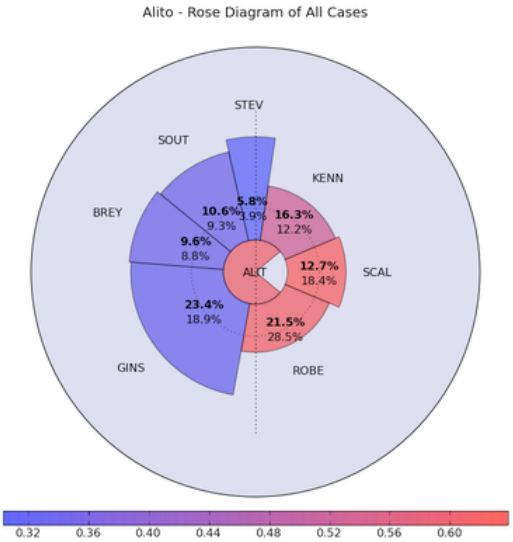
Though sentiment analysis would likely make considerable contributions to the quality of Supreme Court forecasting as suggested by Wrightsman (2008) and Johnson et al. (2009a) automatic detection of sentiment in a domain such as Supreme Court discourse is likely to be considerably harder than the already difficult typical sentiment analysis tasks. While overt sentiment may be expressed by word choice, in a formal setting such as the Supreme Court, sentiment will often not be expressed overtly, thus requiring researchers to rely on methods for identifying covert sentiment (Evans et al. 2007, Green and Resnik 2009). This raises its own issues, as expression of covert sentiment is likely to vary between cases as the issue area of cases changes. These factors make the task of automatic sentiment detection in this domain a considerably different task than typical areas of sentiment detection such as movie and product reviews.

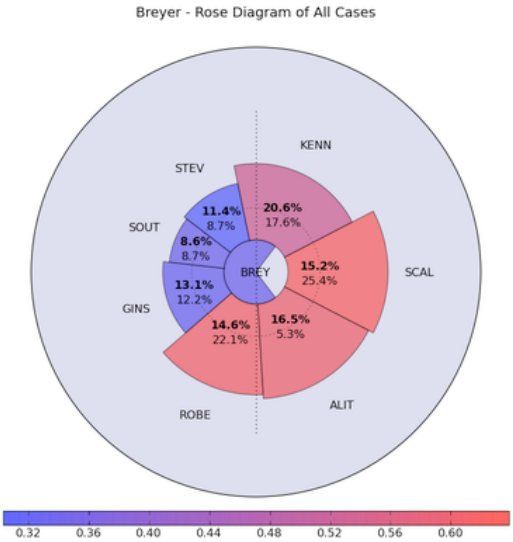
In Chapter 1 we discussed the potential broader implications of this research. That is, this work could be extended to other situations where we are interested in the relationship between conversational behavior and non-linguistic actions. While we are confident that we could directly apply these approaches to other similar situations, e.g. lower courts or even contestant judging on reality shows, this opens up the question of just how far approaches similar to those covered here can be applied. Do individuals in conversational settings take on recognizable natural roles (e.g. leader, "devil's advocate", etc.) that are applicable across numerous situations? If so, would we be able to reduce

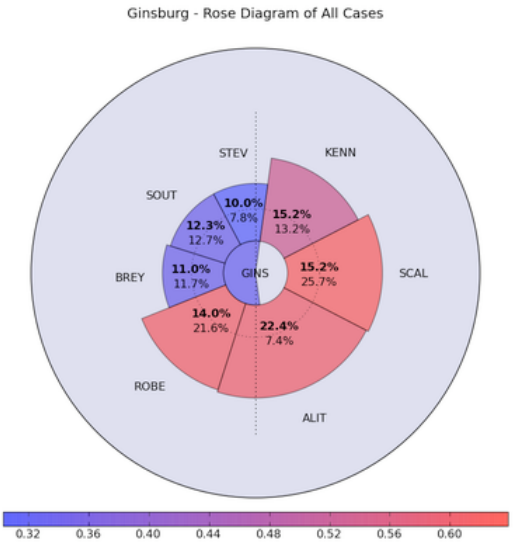
reliance on speaker and domain specific training data, expanding the applicability of these approaches to a wider range of conversational settings such as business negotiations and other meetings? And, what might we learn about human interaction in general and the relationship between conversational interaction and real world actions from these sorts of approaches? By exploring the conversational dynamics of the U.S. Supreme Court and their relationship with the actions taken by the Court as a whole and by individual justices, this work begins to address these questions.

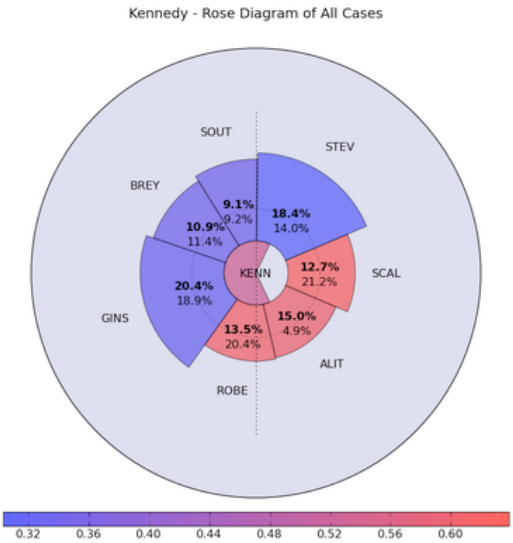
Appendix A Rose Charts

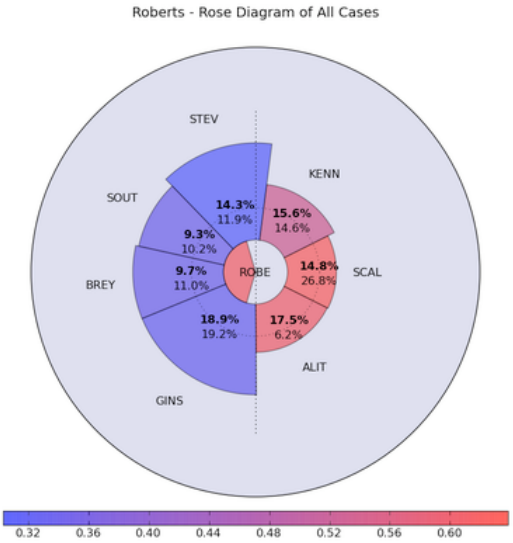
All Cases

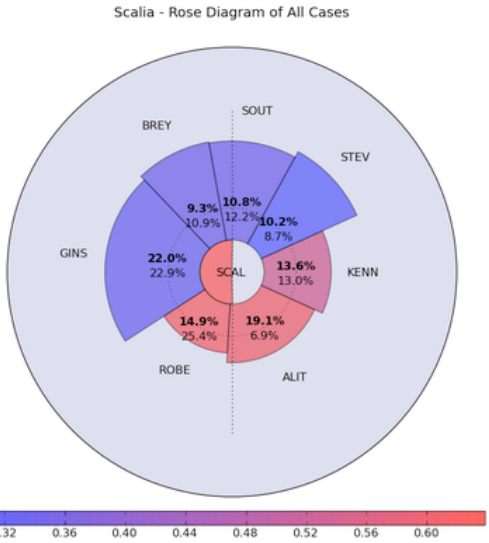


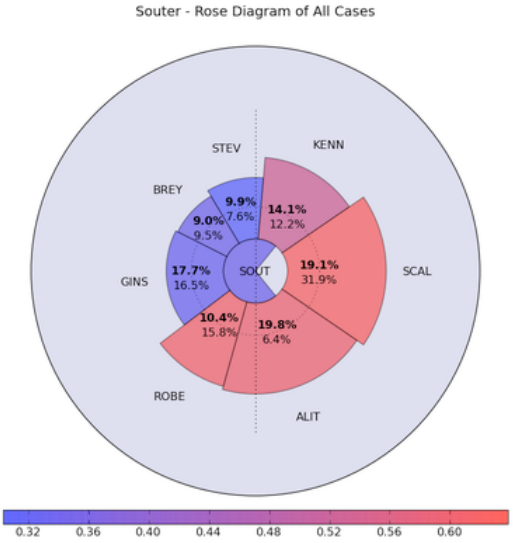


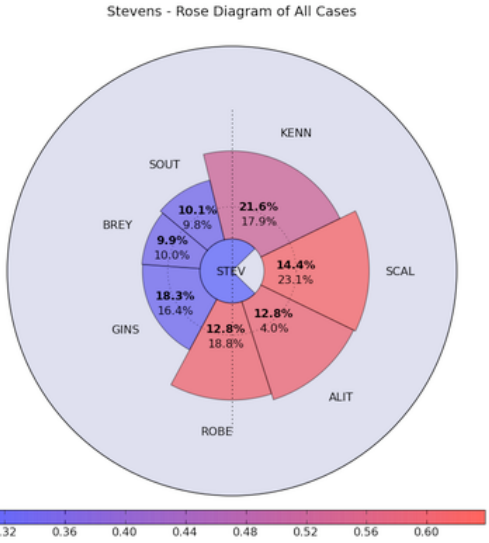




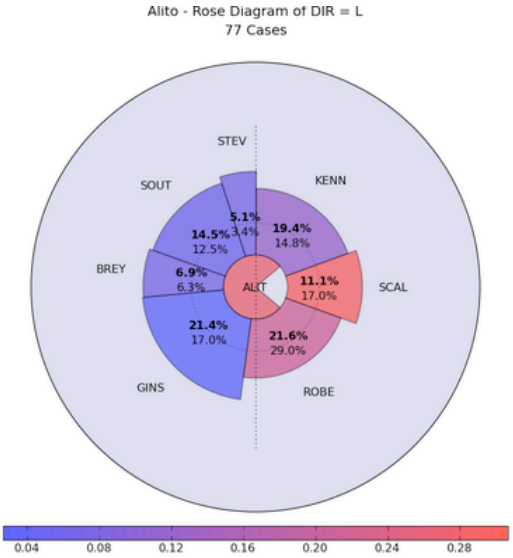

Roberts - Rose Diagram of All Cases
STEV
KENN
SOUT
BREY
ROBE
SCAL
ALIT
GINS
14.3%
11.9%
15.6%
14.6%
9.3%
10.2%
9.7%
11.0%
14.8%
26.8%
18.9%
19.2%
17.5%
6.2%
0.32 0.36 0.40 0.44 0.48 0.52 0.56 0.60

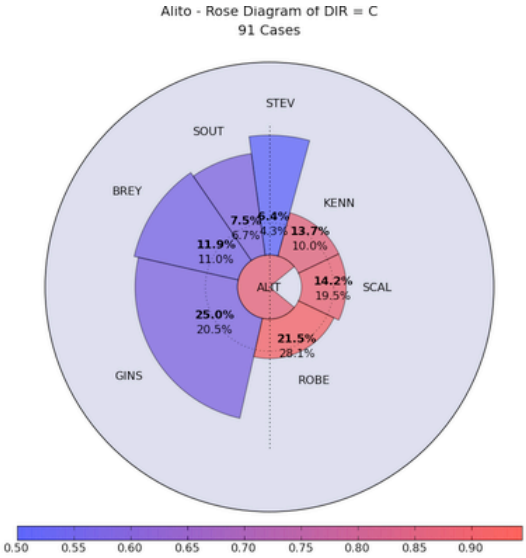

Scalia - Rose Diagram of All Cases
SOUT
BREY
STEV
GINS
SCAL
KENN
ROBE
ALIT
10.8%
12.2%
9.3%
10.9%
10.2%
8.7%
22.0%
22.9%
13.6%
13.0%
14.9%
25.4%
19.1%
6.9%
0.32 0.36 0.40 0.44 0.48 0.52 0.56 0.60

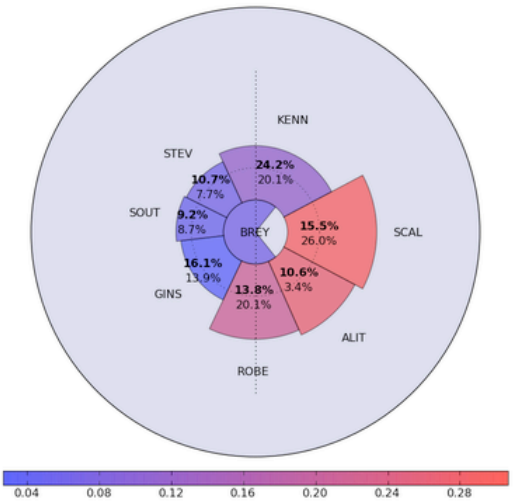

Souter - Rose Diagram of All Cases
STEV
KENN
BREY
GINS
SOUT
SCAL
ROBE
ALIT
9.9%
7.6%
14.1%
12.2%
9.0%
9.5%
17.7%
16.5%
19.1%
31.9%
10.4%
15.8%
19.8%
6.4%
0.32 0.36 0.40 0.44 0.48 0.52 0.56 0.60

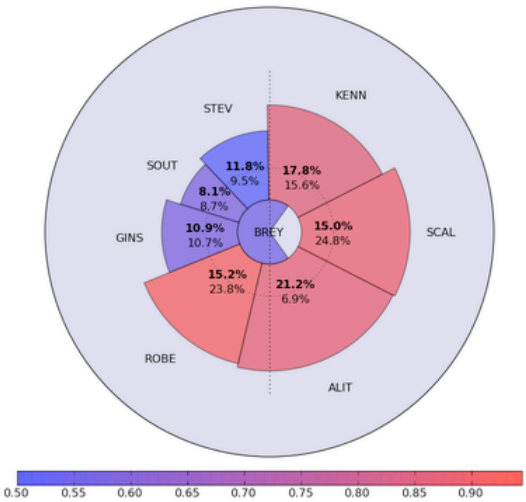

Stevens - Rose Diagram of All Cases
KENN
SOUT
BREY
GINS
STEV
SCAL
ALIT
ROBE
10.1%
9.8%
21.6%
17.9%
9.9%
10.0%
18.3%
16.4%
14.4%
23.1%
12.8%
18.8%
12.8%
4.0%
0.32 0.36 0.40 0.44 0.48 0.52 0.56 0.60

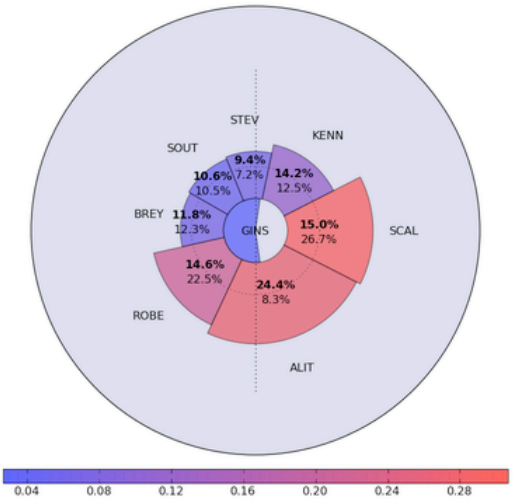
DIR Condition

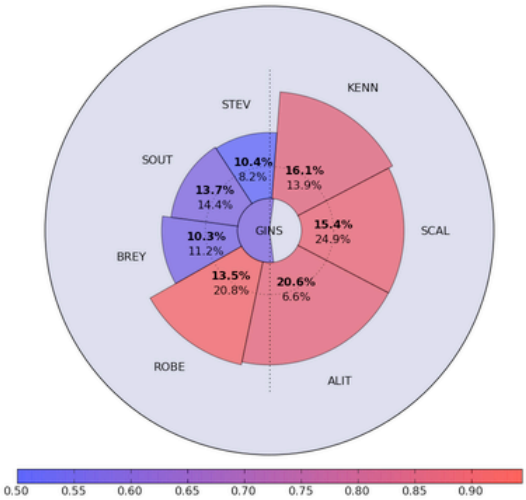


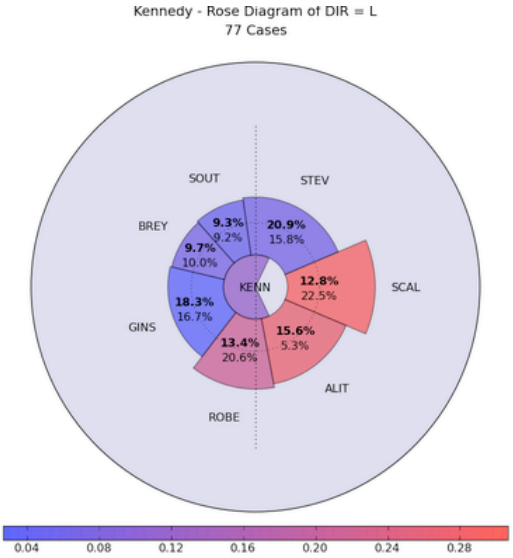


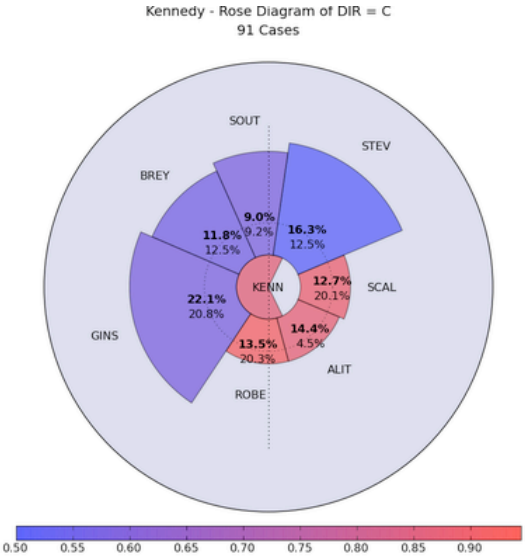


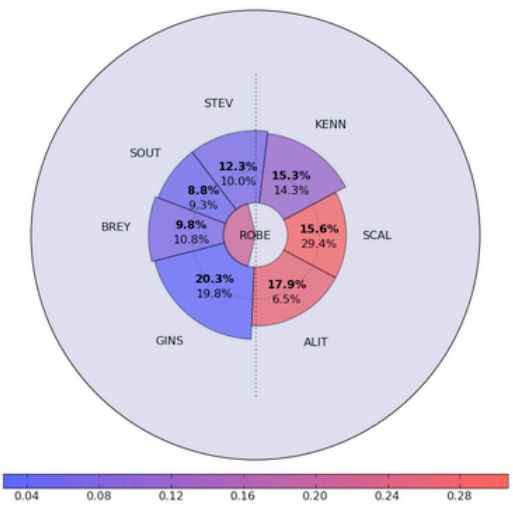


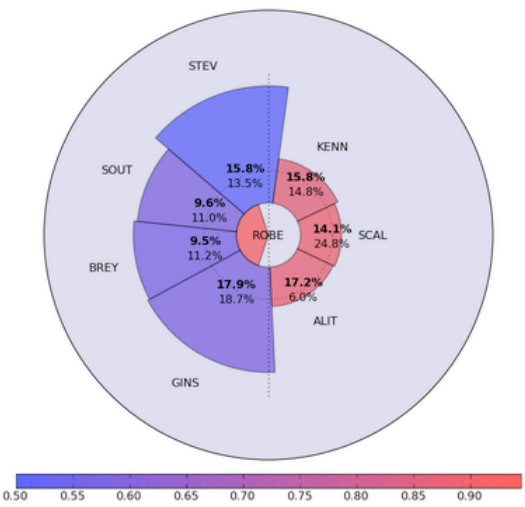


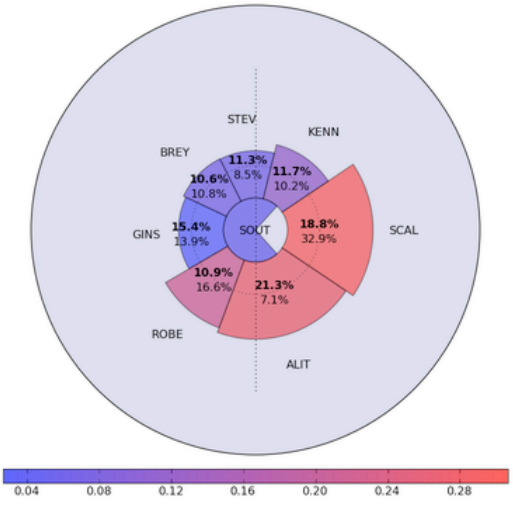


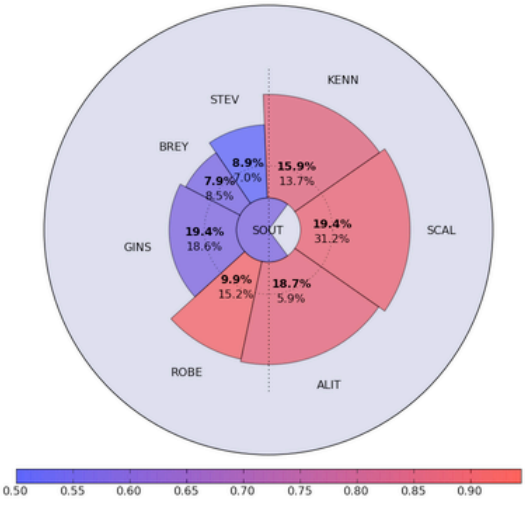

Kennedy - Rose Diagram of DIR = L
77 Cases
SOUT
STEV
BREY
GINS
SCAL
ALIT
ROBE
KENN
9.3%
9.2%
20.9%
15.8%
9.7%
10.0%
12.8%
22.5%
18.3%
16.7%
15.6%
5.3%
13.4%
20.6%
0.04 0.08 0.12 0.16 0.20 0.24 0.28

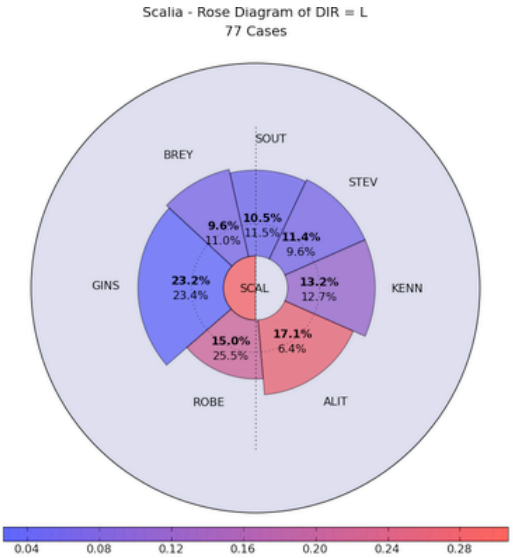

Kennedy - Rose Diagram of DIR = C
91 Cases
SOUT
STEV
BREY
GINS
SCAL
ALIT
ROBE
KENN
9.0%
9.2%
16.3%
12.5%
11.8%
12.5%
12.7%
20.1%
22.1%
20.8%
14.4%
4.5%
13.5%
20.3%
0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90

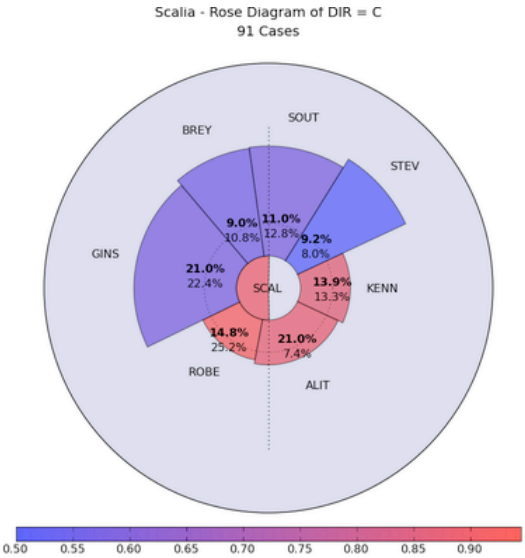

Roberts - Rose Diagram of DIR = L
77 Cases
STEV
KENN
SOUT
BREY
SCAL
GINS
ALIT
ROBE
12.3%
10.0%
15.3%
14.3%
8.8%
9.3%
9.8%
10.8%
15.6%
29.4%
20.3%
19.8%
17.9%
6.5%
0.04 0.08 0.12 0.16 0.20 0.24 0.28

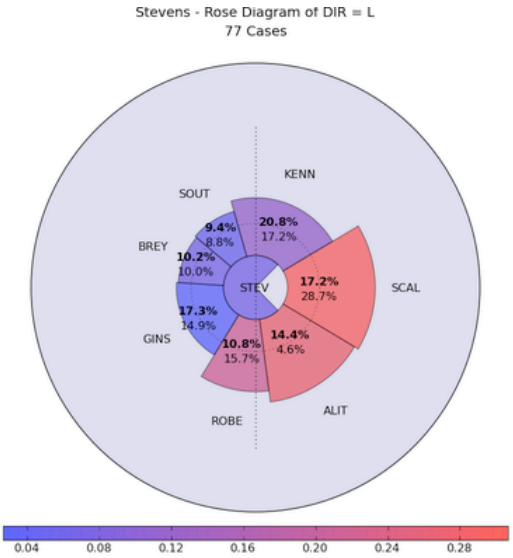

Roberts - Rose Diagram of DIR = C
91 Cases
STEV
KENN
SOUT
BREY
SCAL
GINS
ALIT
ROBE
15.8%
13.5%
15.8%
14.8%
9.6%
11.0%
14.1%
24.8%
9.5%
11.2%
17.9%
18.7%
17.2%
6.0%
0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90

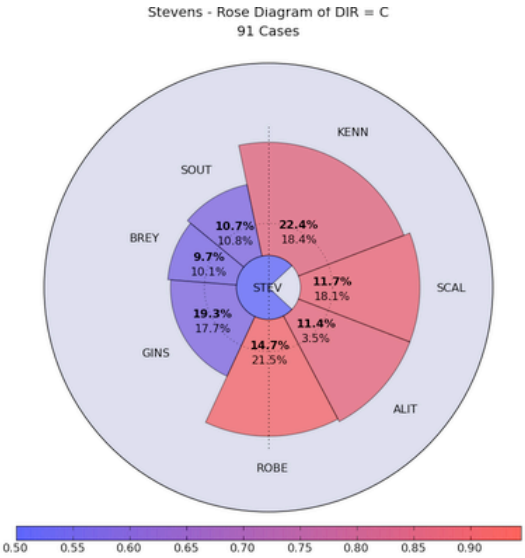

Souter - Rose Diagram of DIR = L
77 Cases
STEV
KENN
BREY
GINS
SCAL
ALIT
ROBE
SOUT
11.3%
8.5%
11.7%
10.2%
10.6%
10.8%
18.8%
32.9%
15.4%
13.9%
10.9%
16.6%
21.3%
7.1%
0.04 0.08 0.12 0.16 0.20 0.24 0.28


Souter - Rose Diagram of DIR = C
91 Cases
KENN
STEV
BREY
GINS
SCAL
ALIT
ROBE
SOUT
8.9%
7.0%
15.9%
13.7%
7.9%
8.5%
19.4%
31.2%
19.4%
18.6%
9.9%
15.2%
18.7%
5.9%
0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90

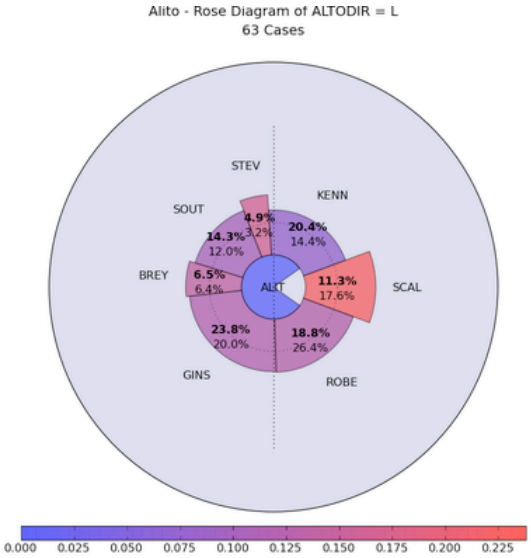

Scalia - Rose Diagram of DIR = L
77 Cases
SOUT
BREY
STEV
GINS
KENN
SCAL
ROBE
ALIT
9.6%
11.0%
10.5%
11.5%
11.4%
9.6%
23.2%
23.4%
13.2%
12.7%
15.0%
25.5%
17.1%
6.4%
0.04
0.08
0.12
0.16
0.20
0.24
0.28

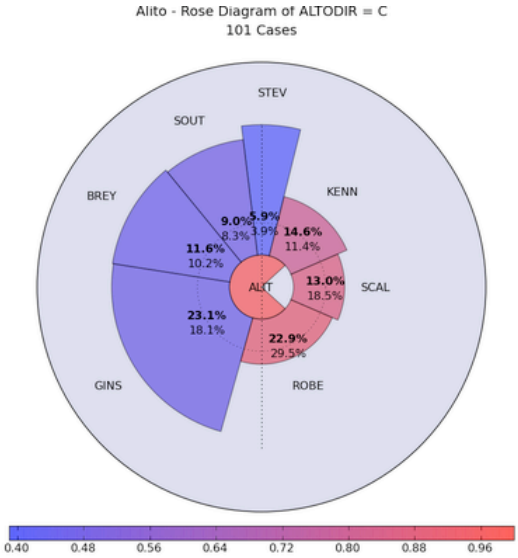

Scalia - Rose Diagram of DIR = C
91 Cases
SOUT
BREY
STEV
GINS
KENN
SCAL
ROBE
ALIT
9.0%
10.8%
11.0%
12.8%
9.2%
8.0%
21.0%
22.4%
13.9%
13.3%
14.8%
25.2%
21.0%
7.4%
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90

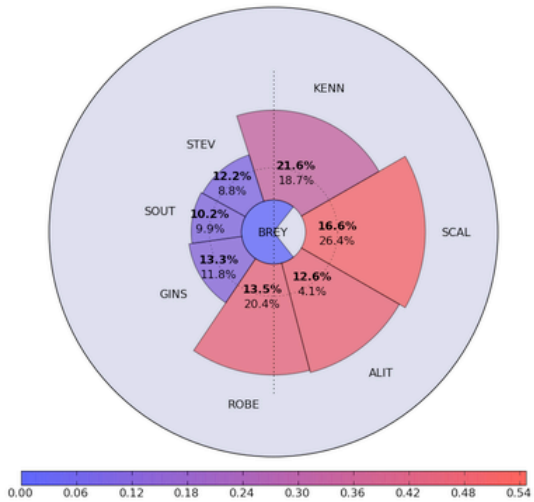

Stevens - Rose Diagram of DIR = L
77 Cases
KENN
SOUT
BREY
STEV
SCAL
GINS
ROBE
ALIT
9.4%
8.8%
20.8%
17.2%
10.2%
10.0%
17.2%
28.7%
17.3%
14.9%
10.8%
15.7%
14.4%
4.6%
0.04
0.08
0.12
0.16
0.20
0.24
0.28

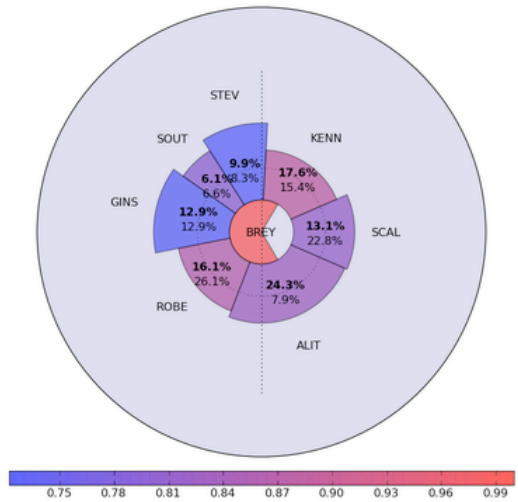

Stevens - Rose Diagram of DIR = C
91 Cases
KENN
SOUT
BREY
STEV
SCAL
GINS
ROBE
ALIT
10.7%
10.8%
22.4%
18.4%
9.7%
10.1%
11.7%
18.1%
19.3%
17.7%
11.4%
3.5%
14.7%
21.5%
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90

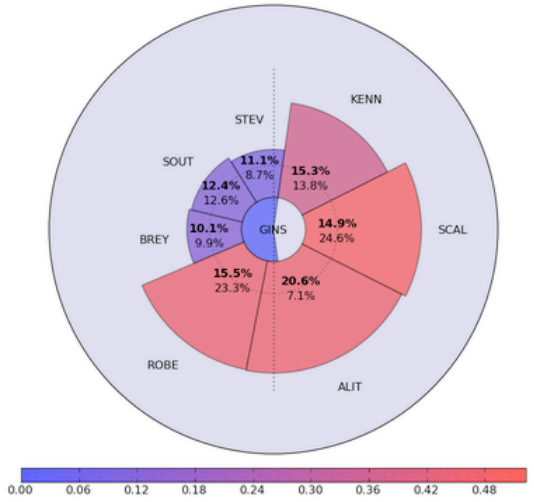
JDIR Condition

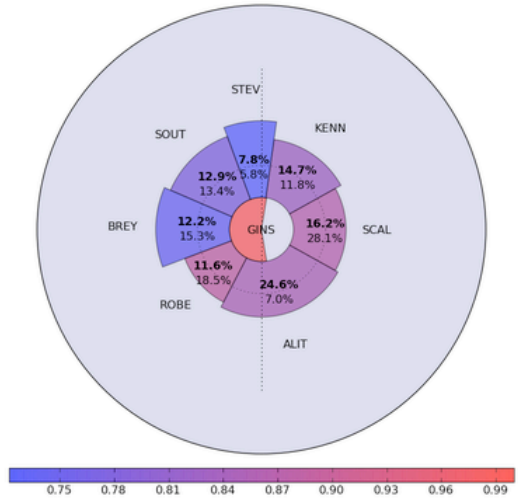




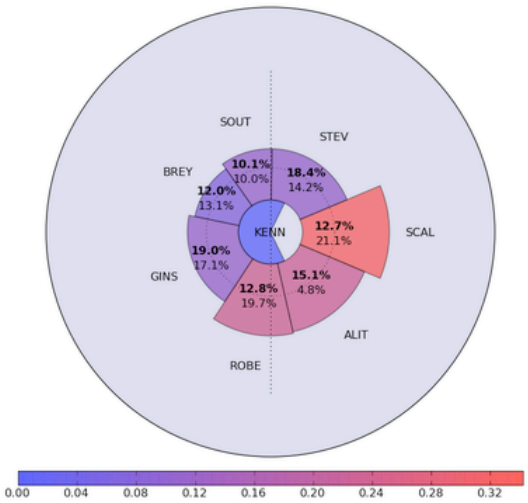




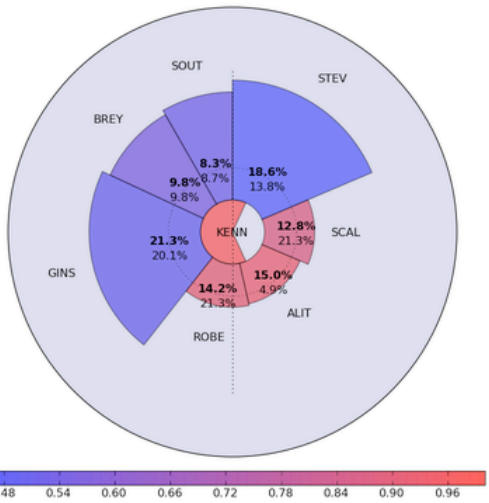




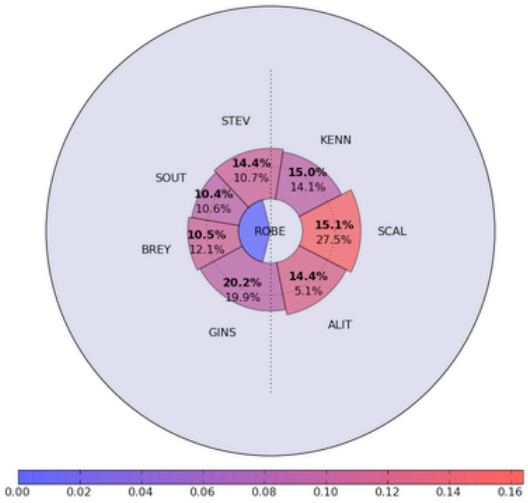
Kennedy - Rose Diagram of KENDIR = L
76 Cases



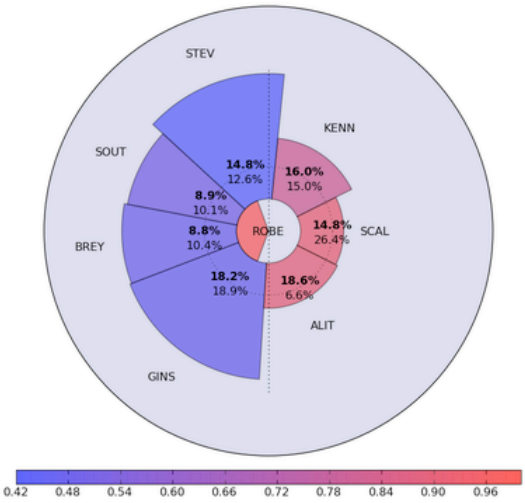
Kennedy - Rose Diagram of KENDIR = C
89 Cases



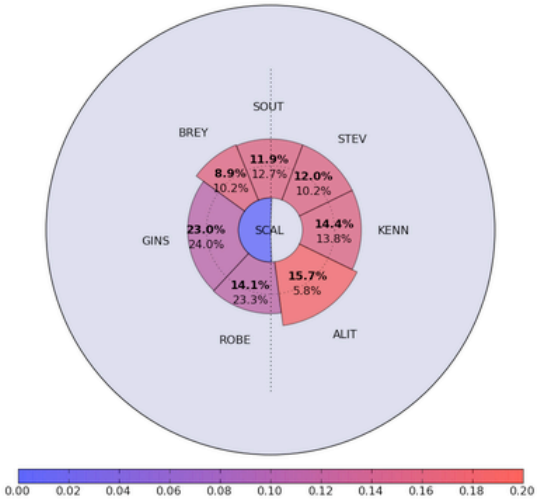
Roberts - Rose Diagram of ROBTDIR = L
61 Cases



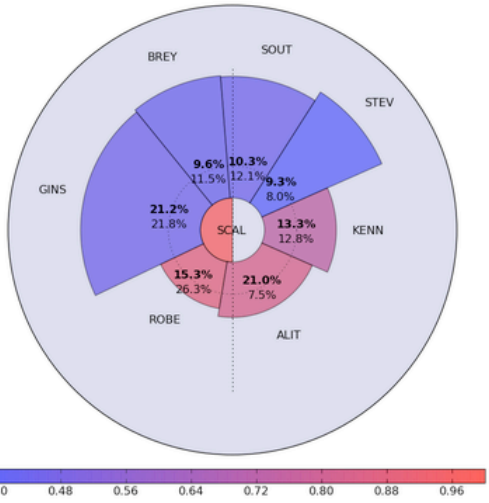
Roberts - Rose Diagram of ROBTDIR = C
101 Cases

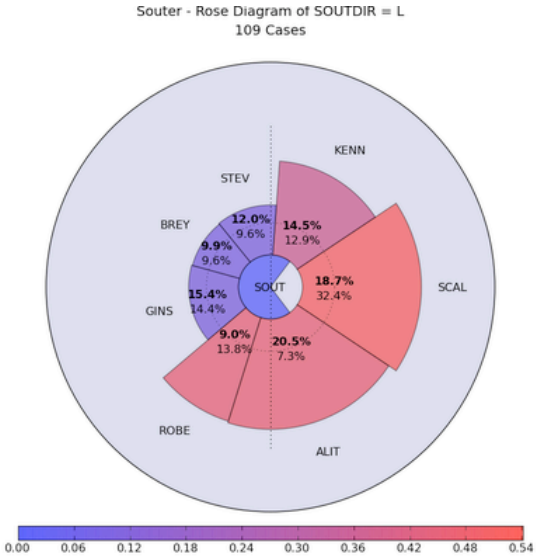


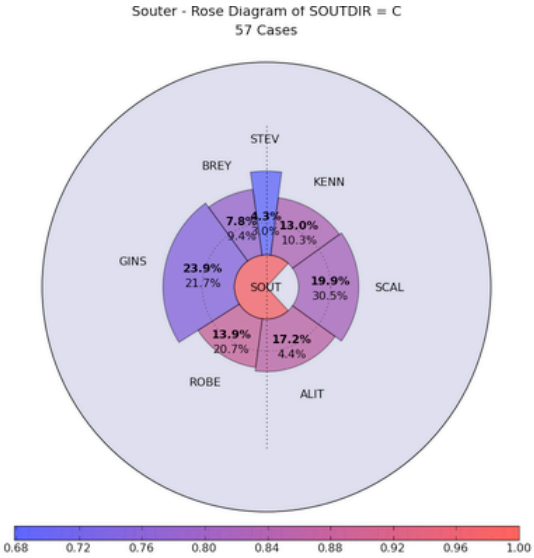
Scalia - Rose Diagram of SCALDIR = L
60 Cases

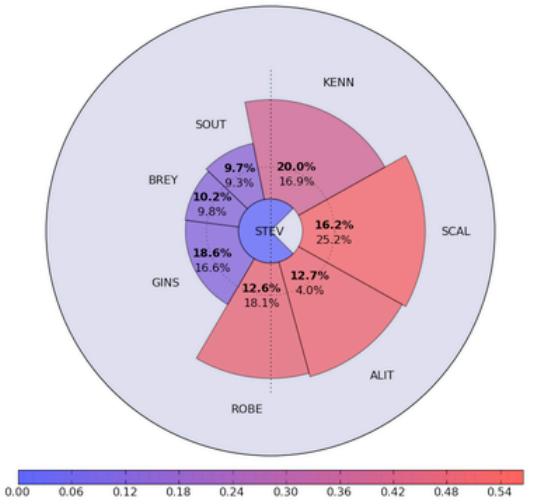


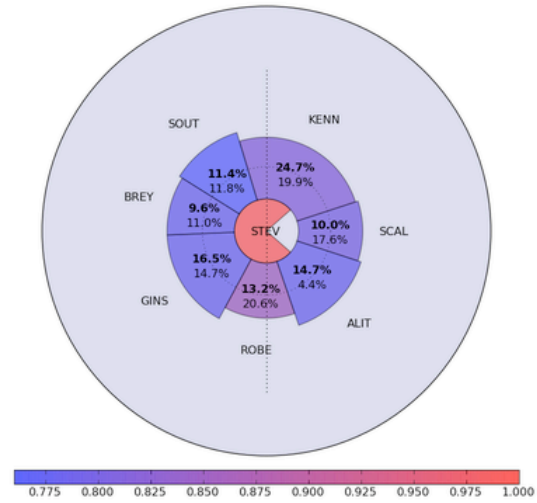
Scalia - Rose Diagram of SCALDIR = C
106 Cases



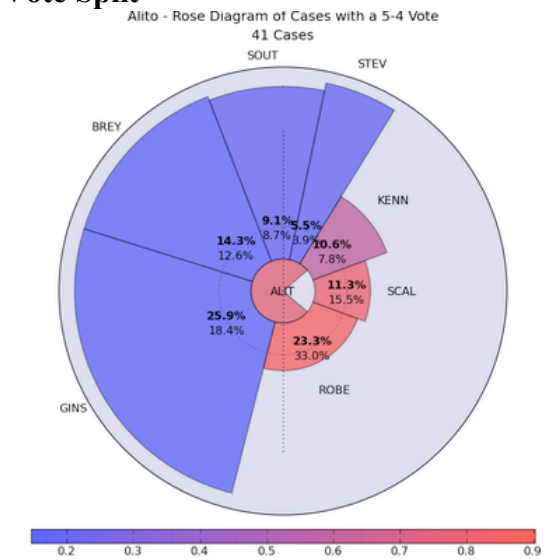

Souter - Rose Diagram of SOUTDIR = L
109 Cases
KENN
STEV
BREY
SCAL
GINS
ROBE
ALIT
SOUT
12.0%
9.6%
14.5%
12.9%
9.9%
9.6%
18.7%
32.4%
15.4%
14.4%
9.0%
13.8%
20.5%
7.3%
0.00 0.06 0.12 0.18 0.24 0.30 0.36 0.42 0.48 0.54

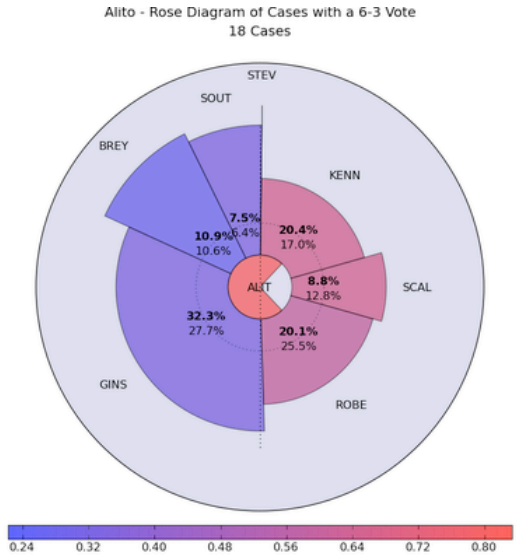

Souter - Rose Diagram of SOUTDIR = C
57 Cases
STEV
BREY
KENN
GINS
SCAL
ROBE
ALIT
SOUT
7.8%
9.4%
13.0%
10.3%
23.9%
21.7%
19.9%
30.5%
13.9%
20.7%
17.2%
4.4%
0.68 0.72 0.76 0.80 0.84 0.88 0.92 0.96 1.00

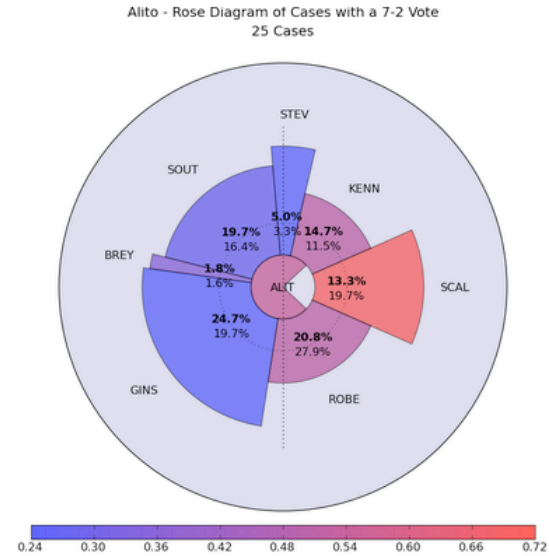

Stevens - Rose Diagram of STEVDIR = L
115 Cases
KENN
SOUT
BREY
SCAL
GINS
ALIT
ROBE
STEV
9.7%
9.3%
20.0%
16.9%
10.2%
9.8%
16.2%
25.2%
18.6%
16.6%
12.6%
18.1%
12.7%
4.0%
0.00 0.06 0.12 0.18 0.24 0.30 0.36 0.42 0.48 0.54

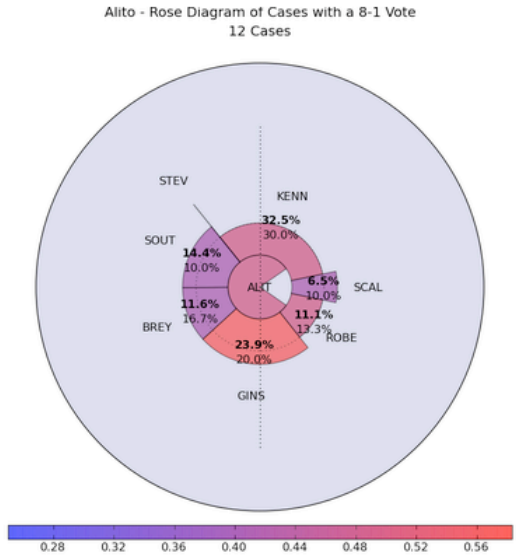

Stevens - Rose Diagram of STEVDIR = C
50 Cases
SOUT
KENN
BREY
SCAL
GINS
ALIT
ROBE
STEV
11.4%
11.8%
24.7%
19.9%
9.6%
11.0%
10.0%
17.6%
16.5%
14.7%
13.2%
20.6%
14.7%
4.4%
0.775 0.800 0.825 0.850 0.875 0.900 0.925 0.950 0.975 1.000

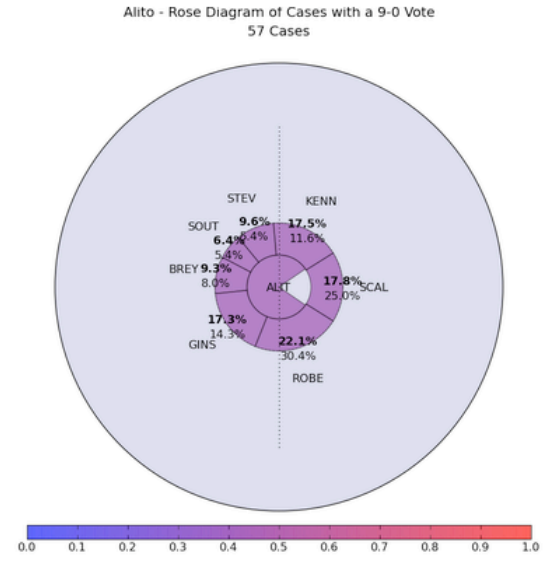
Vote Split

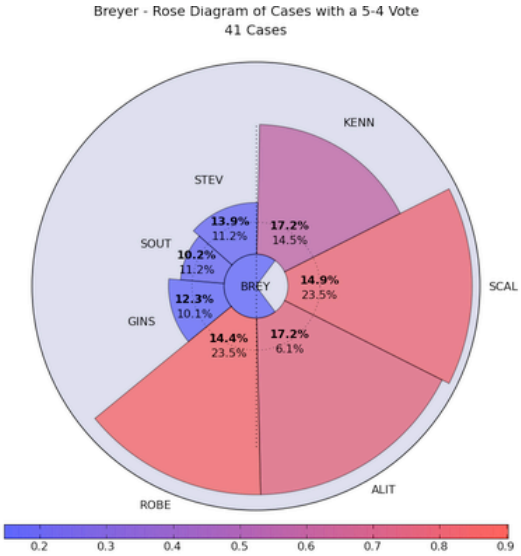

Alito - Rose Diagram of Cases with a 5-4 Vote
41 Cases
SOUT
STEV
BREY
KENN
SCAL
ROBE
GINS
ALIT
9.1%
8.7%
5.5%
3.9%
14.3%
12.6%
10.6%
7.8%
11.3%
15.5%
25.9%
18.4%
23.3%
33.0%
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

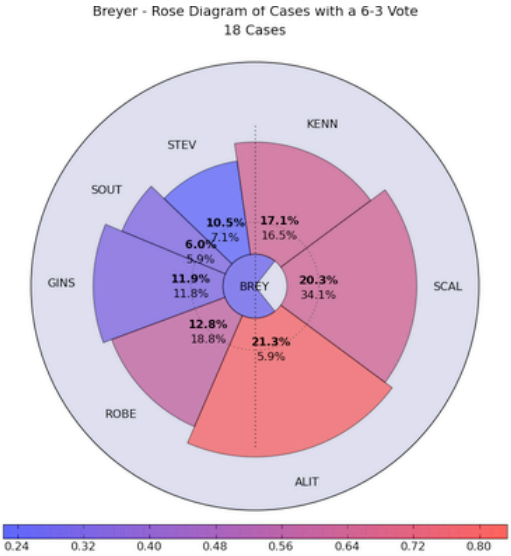

Alito - Rose Diagram of Cases with a 6-3 Vote
18 Cases
STEV
SOUT
BREY
KENN
SCAL
ROBE
GINS
ALIT
7.5%
5.4%
20.4%
17.0%
10.9%
10.6%
8.8%
12.8%
32.3%
27.7%
20.1%
25.5%
0.24 0.32 0.40 0.48 0.56 0.64 0.72 0.80

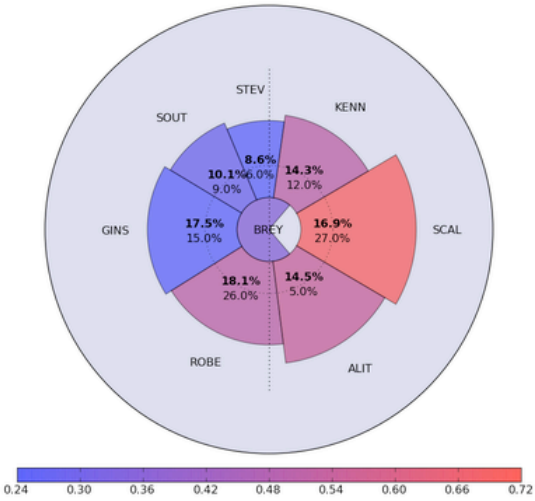

Alito - Rose Diagram of Cases with a 7-2 Vote
25 Cases
STEV
SOUT
KENN
BREY
SCAL
GINS
ROBE
ALIT
5.0%
3.3%
14.7%
11.5%
19.7%
16.4%
1.8%
1.6%
13.3%
19.7%
24.7%
19.7%
20.8%
27.9%
0.24 0.30 0.36 0.42 0.48 0.54 0.60 0.66 0.72

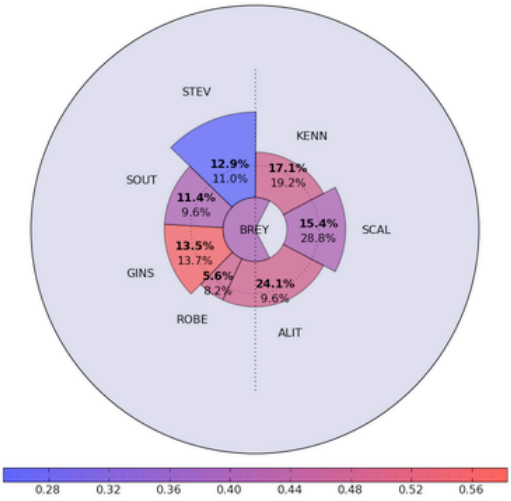

Alito - Rose Diagram of Cases with a 8-1 Vote
12 Cases
STEV
KENN
SOUT
SCAL
BREY
ROBE
GINS
ALIT
32.5%
30.0%
14.4%
10.0%
6.5%
10.0%
11.6%
16.7%
11.1%
13.3%
23.9%
20.0%
0.28 0.32 0.36 0.40 0.44 0.48 0.52 0.56

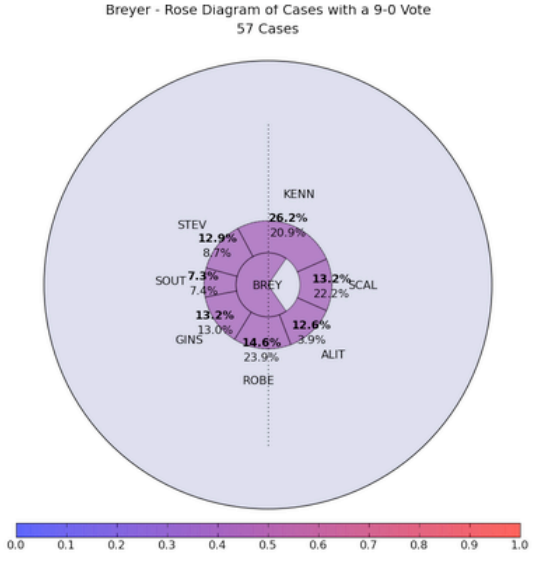

Alito - Rose Diagram of Cases with a 9-0 Vote
57 Cases
STEV
KENN
SOUT
SCAL
BREY
GINS
ROBE
ALIT
9.6%
6.4%
5.4%
17.5%
11.6%
9.3%
8.0%
17.8%
25.0%
17.3%
14.3%
22.1%
30.4%
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

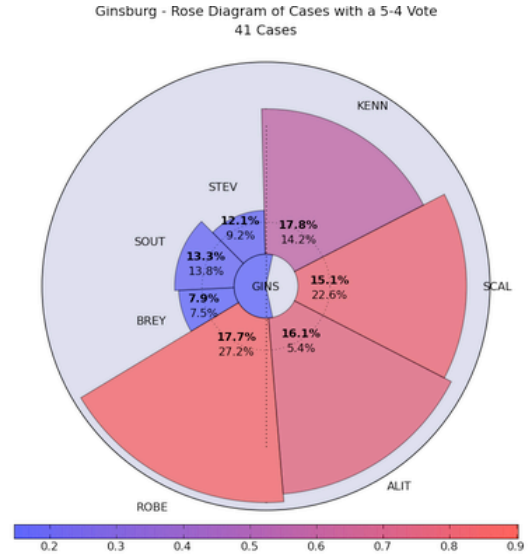

Breyer - Rose Diagram of Cases with a 5-4 Vote
41 Cases
KENN
SCAL
ALIT
ROBE
GINS
SOUT
STEV
BREY
17.2%
14.5%
14.9%
23.5%
17.2%
6.1%
14.4%
23.5%
12.3%
10.1%
10.2%
11.2%
13.9%
11.2%
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

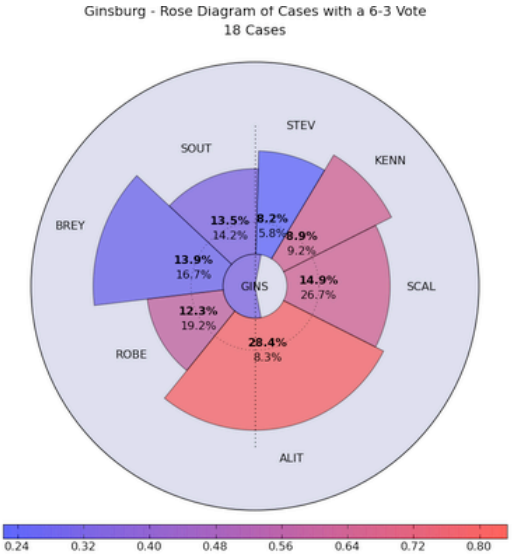

Breyer - Rose Diagram of Cases with a 6-3 Vote
18 Cases
KENN
SCAL
ALIT
ROBE
GINS
SOUT
STEV
BREY
17.1%
16.5%
20.3%
34.1%
21.3%
5.9%
12.8%
18.8%
11.9%
11.8%
6.0%
5.9%
10.5%
7.1%
0.24 0.32 0.40 0.48 0.56 0.64 0.72 0.80

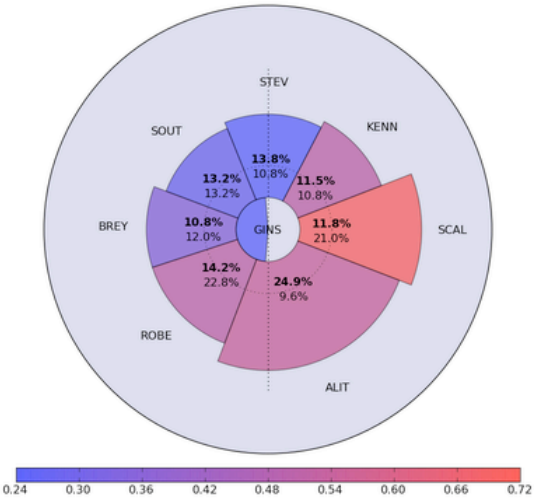

Breyer - Rose Diagram of Cases with a 7-2 Vote
25 Cases
KENN
SCAL
ALIT
ROBE
GINS
SOUT
STEV
BREY
14.3%
12.0%
16.9%
27.0%
14.5%
5.0%
18.1%
26.0%
17.5%
15.0%
10.1%
9.0%
8.6%
6.0%
0.24 0.30 0.36 0.42 0.48 0.54 0.60 0.66 0.72

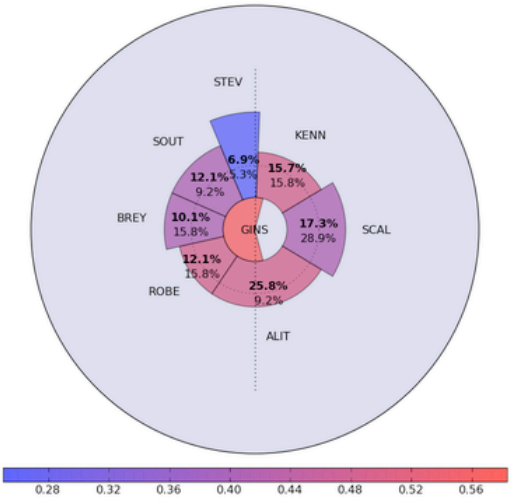

Breyer - Rose Diagram of Cases with a 8-1 Vote
12 Cases
KENN
SCAL
ALIT
ROBE
GINS
SOUT
STEV
BREY
17.1%
19.2%
15.4%
28.8%
24.1%
9.6%
5.6%
8.2%
13.5%
13.7%
11.4%
9.6%
12.9%
11.0%
0.28 0.32 0.36 0.40 0.44 0.48 0.52 0.56

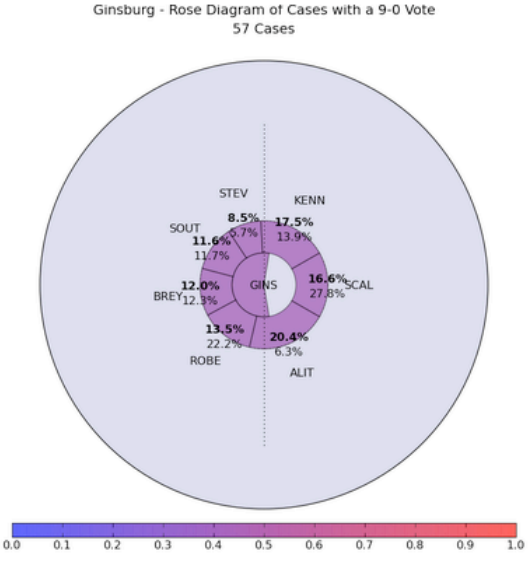

Breyer - Rose Diagram of Cases with a 9-0 Vote
57 Cases
KENN
SCAL
ALIT
ROBE
GINS
SOUT
STEV
BREY
26.2%
20.9%
13.2%
22.2%
12.6%
3.9%
14.6%
23.9%
13.2%
13.0%
7.3%
7.4%
12.9%
8.7%
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

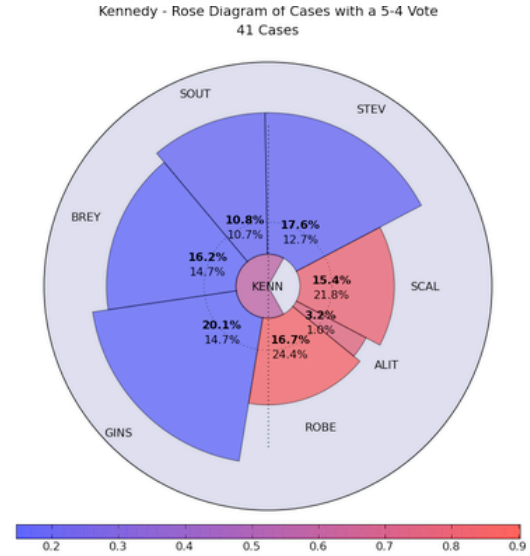

Ginsburg - Rose Diagram of Cases with a 5-4 Vote
41 Cases
KENN
STEV
SOUT
SCAL
BREY
GINS
ROBE
ALIT
12.1%
9.2%
17.8%
14.2%
13.3%
13.8%
15.1%
22.6%
7.9%
7.5%
17.7%
27.2%
16.1%
5.4%
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

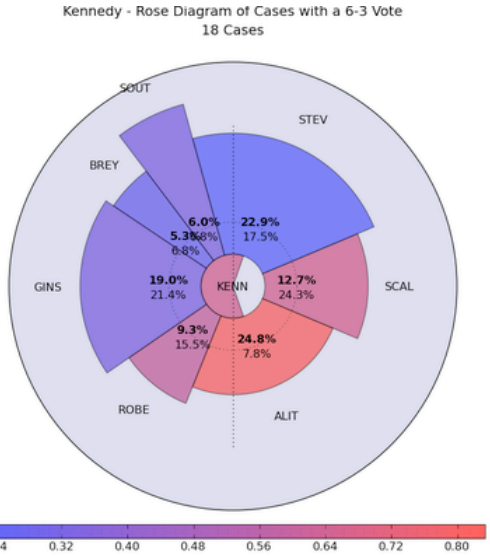

Ginsburg - Rose Diagram of Cases with a 6-3 Vote
18 Cases
STEV
SOUT
KENN
BREY
SCAL
GINS
ROBE
ALIT
13.5%
14.2%
8.2%
5.8%
8.9%
9.2%
13.9%
16.7%
14.9%
26.7%
12.3%
19.2%
28.4%
8.3%
0.24
0.32
0.40
0.48
0.56
0.64
0.72
0.80

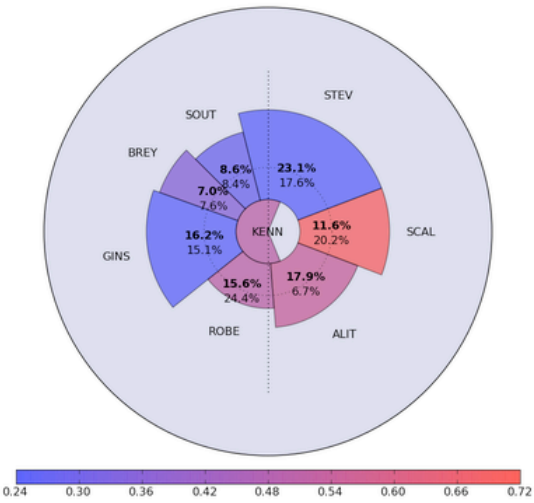

Ginsburg - Rose Diagram of Cases with a 7-2 Vote
25 Cases
STEV
SOUT
KENN
BREY
SCAL
GINS
ROBE
ALIT
13.8%
10.8%
13.2%
13.2%
11.5%
10.8%
10.8%
12.0%
11.8%
21.0%
14.2%
22.8%
24.9%
9.6%
0.24
0.30
0.36
0.42
0.48
0.54
0.60
0.66
0.72

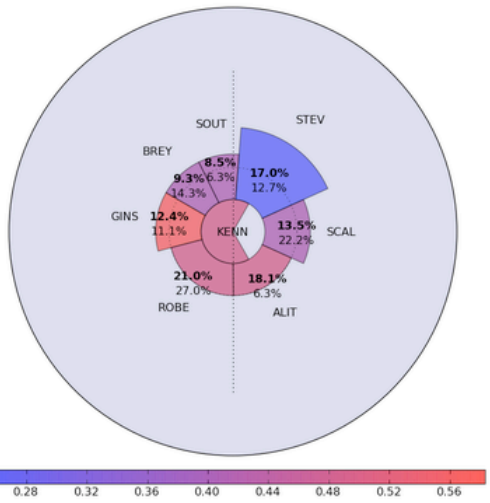

Ginsburg - Rose Diagram of Cases with a 8-1 Vote
12 Cases
STEV
SOUT
KENN
BREY
SCAL
GINS
ROBE
ALIT
6.9%
5.3%
15.7%
15.8%
12.1%
9.2%
10.1%
15.8%
17.3%
28.9%
12.1%
15.8%
25.8%
9.2%
0.28
0.32
0.36
0.40
0.44
0.48
0.52
0.56

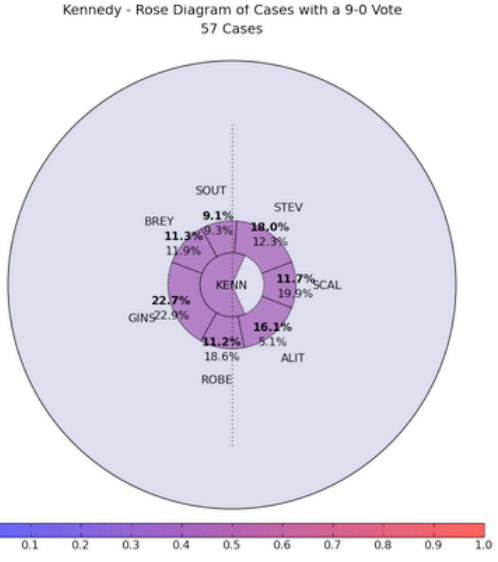

Ginsburg - Rose Diagram of Cases with a 9-0 Vote
57 Cases
STEV
KENN
SOUT
SCAL
BREY
GINS
ROBE
ALIT
8.5%
8.7%
17.5%
13.9%
11.6%
11.7%
16.6%
27.8%
12.0%
12.3%
13.5%
22.2%
20.4%
6.3%
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

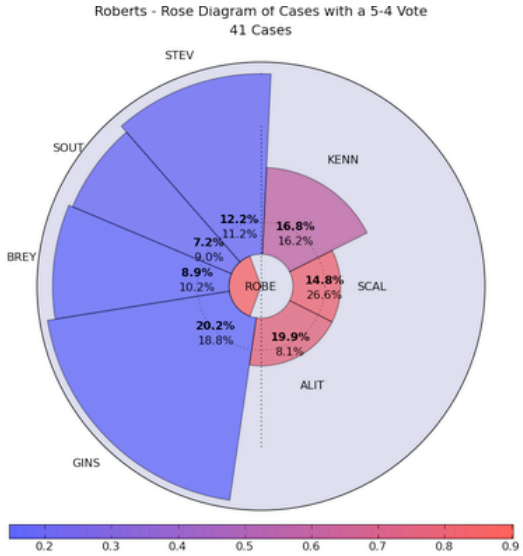

Kennedy - Rose Diagram of Cases with a 5-4 Vote
41 Cases
SOUT
STEV
BREY
SCAL
GINS
ALIT
ROBE
KENN
10.8%
10.7%
17.6%
12.7%
16.2%
14.7%
15.4%
21.8%
3.2%
1.6%
20.1%
14.7%
16.7%
24.4%
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

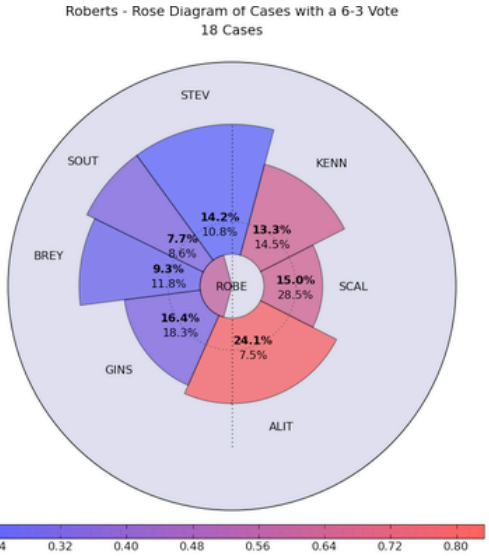

Kennedy - Rose Diagram of Cases with a 6-3 Vote
18 Cases
SOUT
STEV
BREY
GINS
SCAL
ROBE
ALIT
KENN
6.0%
5.3%
6.8%
22.9%
17.5%
19.0%
21.4%
12.7%
24.3%
9.3%
15.5%
24.8%
7.8%
0.24 0.32 0.40 0.48 0.56 0.64 0.72 0.80

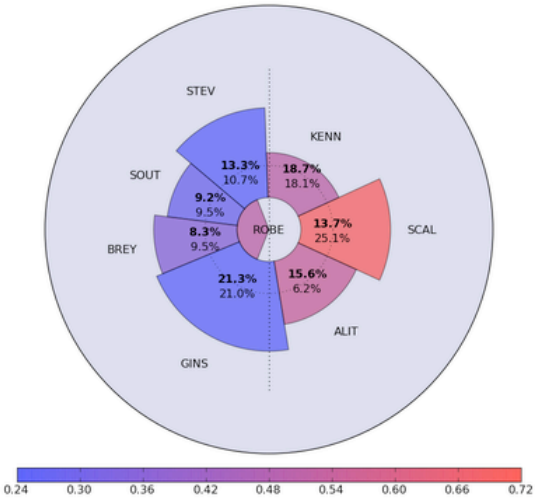

Kennedy - Rose Diagram of Cases with a 7-2 Vote
25 Cases
STEV
SOUT
BREY
GINS
SCAL
ROBE
ALIT
KENN
8.6%
8.4%
23.1%
17.6%
7.0%
7.6%
16.2%
15.1%
11.6%
20.2%
15.6%
24.4%
17.9%
6.7%
0.24 0.30 0.36 0.42 0.48 0.54 0.60 0.66 0.72

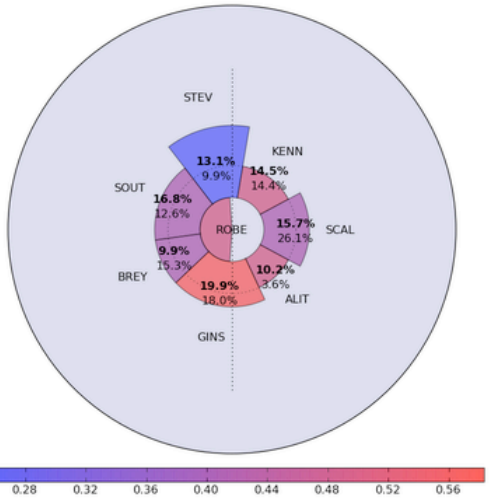

Kennedy - Rose Diagram of Cases with a 8-1 Vote
12 Cases
SOUT
STEV
BREY
GINS
SCAL
ROBE
ALIT
KENN
8.5%
6.3%
17.0%
12.7%
9.3%
14.3%
12.4%
11.1%
13.5%
22.2%
21.0%
27.0%
18.1%
6.3%
0.28 0.32 0.36 0.40 0.44 0.48 0.52 0.56

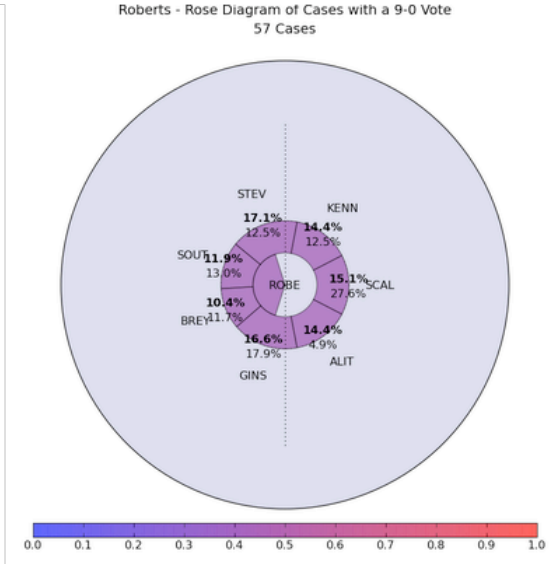

Kennedy - Rose Diagram of Cases with a 9-0 Vote
57 Cases
SOUT
STEV
BREY
SCAL
GINS
ALIT
ROBE
KENN
9.1%
9.3%
18.0%
12.3%
11.3%
11.9%
11.7%
19.9%
22.7%
22.9%
16.1%
5.1%
11.2%
18.6%
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

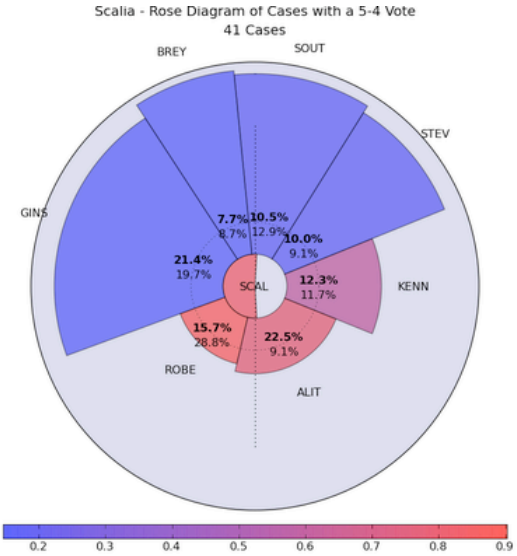

Roberts - Rose Diagram of Cases with a 5-4 Vote
41 Cases
STEV
KENN
SOUT
BREY
ROBE
SCAL
ALIT
GINS
12.2%
11.2%
16.8%
16.2%
7.2%
9.0%
8.9%
10.2%
14.8%
26.6%
20.2%
18.8%
19.9%
8.1%
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

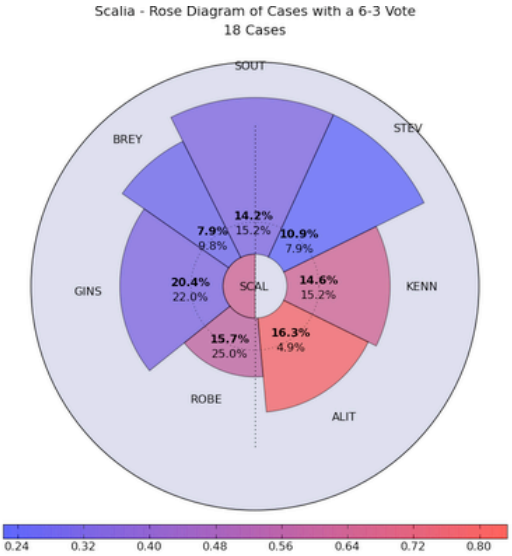

Roberts - Rose Diagram of Cases with a 6-3 Vote
18 Cases
STEV
KENN
SOUT
BREY
ROBE
SCAL
GINS
ALIT
14.2%
10.8%
13.3%
14.5%
7.7%
8.6%
9.3%
11.8%
15.0%
28.5%
16.4%
18.3%
24.1%
7.5%
0.24 0.32 0.40 0.48 0.56 0.64 0.72 0.80

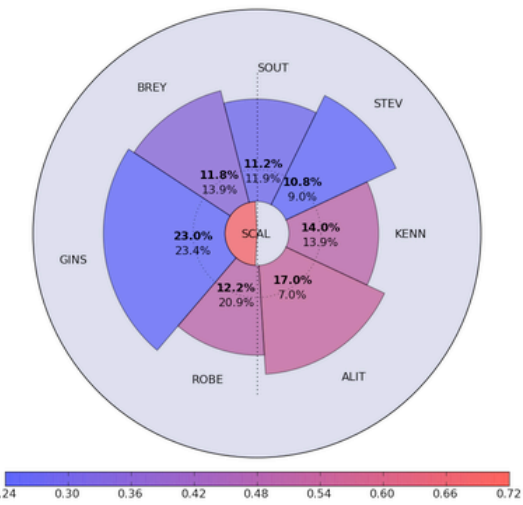

Roberts - Rose Diagram of Cases with a 7-2 Vote
25 Cases
STEV
KENN
SOUT
BREY
ROBE
SCAL
ALIT
GINS
13.3%
10.7%
18.7%
18.1%
9.2%
9.5%
8.3%
9.5%
13.7%
25.1%
21.3%
21.0%
15.6%
6.2%
0.24 0.30 0.36 0.42 0.48 0.54 0.60 0.66 0.72

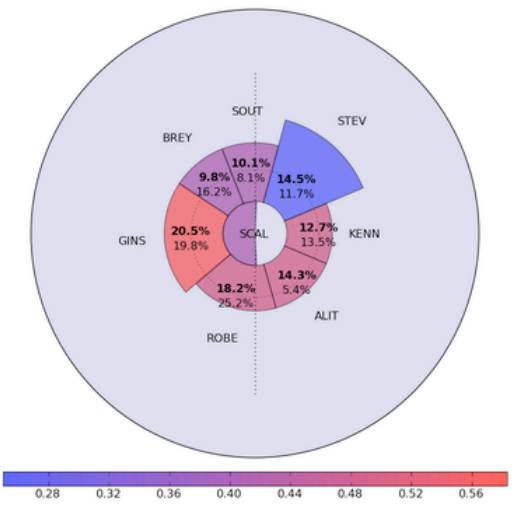

Roberts - Rose Diagram of Cases with a 8-1 Vote
12 Cases
STEV
KENN
SOUT
BREY
ROBE
SCAL
ALIT
GINS
13.1%
9.9%
14.5%
14.4%
16.8%
12.6%
9.9%
15.3%
15.7%
26.1%
19.9%
18.0%
10.2%
3.6%
0.28 0.32 0.36 0.40 0.44 0.48 0.52 0.56

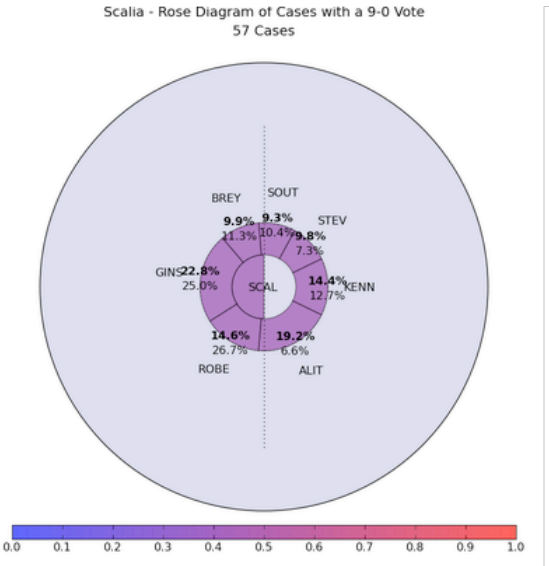

Roberts - Rose Diagram of Cases with a 9-0 Vote
57 Cases
STEV
KENN
SOUT
BREY
ROBE
SCAL
ALIT
GINS
17.1%
12.5%
14.4%
12.5%
11.9%
13.0%
15.1%
27.6%
10.4%
11.7%
16.6%
17.9%
14.4%
4.9%
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

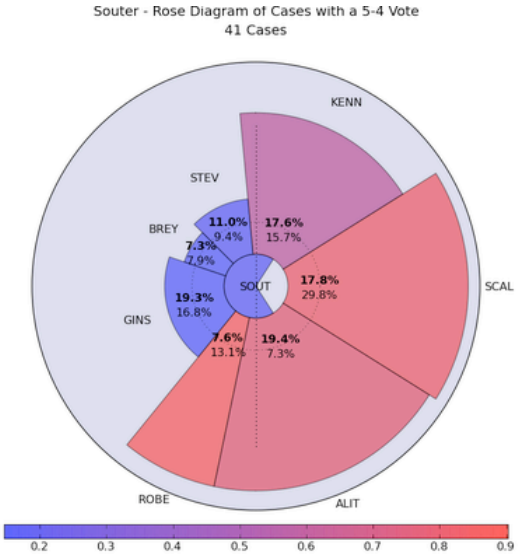

Scalia - Rose Diagram of Cases with a 5-4 Vote
41 Cases
BREY
SOUT
STEV
GINS
KENN
SCAL
ROBE
ALIT
7.7%
8.7%
10.5%
12.9%
10.0%
9.1%
21.4%
19.7%
12.3%
11.7%
15.7%
28.8%
22.5%
9.1%
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

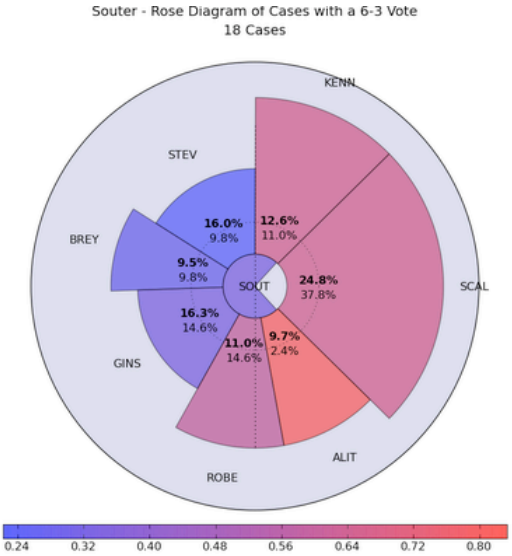

Scalia - Rose Diagram of Cases with a 6-3 Vote
18 Cases
SOUT
BREY
STEV
GINS
KENN
SCAL
ROBE
ALIT
14.2%
15.2%
7.9%
9.8%
10.9%
7.9%
20.4%
22.0%
14.6%
15.2%
15.7%
25.0%
16.3%
4.9%
0.24
0.32
0.40
0.48
0.56
0.64
0.72
0.80

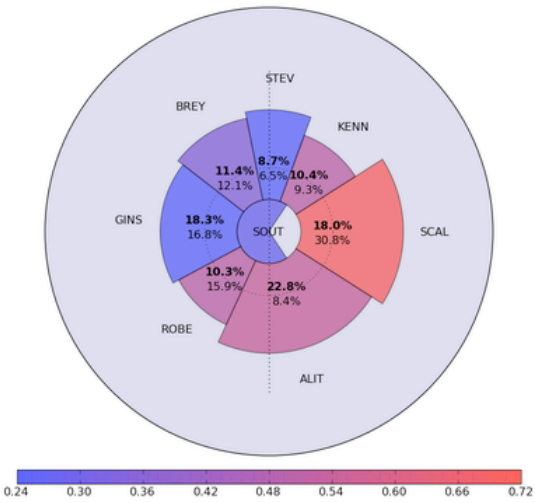

Scalia - Rose Diagram of Cases with a 7-2 Vote
25 Cases
SOUT
BREY
STEV
GINS
KENN
SCAL
ROBE
ALIT
11.2%
11.9%
11.8%
13.9%
10.8%
9.0%
23.0%
23.4%
14.0%
13.9%
12.2%
20.9%
17.0%
7.0%
0.24
0.30
0.36
0.42
0.48
0.54
0.60
0.66
0.72

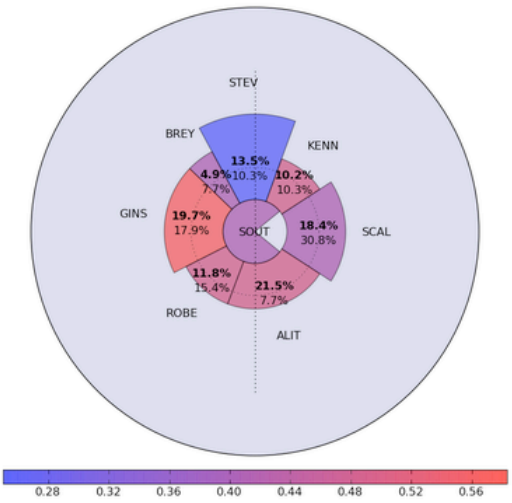

Scalia - Rose Diagram of Cases with a 8-1 Vote
12 Cases
SOUT
BREY
STEV
GINS
KENN
SCAL
ROBE
ALIT
10.1%
8.1%
9.8%
16.2%
14.5%
11.7%
20.5%
19.8%
12.7%
13.5%
18.2%
25.2%
14.3%
5.4%
0.28
0.32
0.36
0.40
0.44
0.48
0.52
0.56

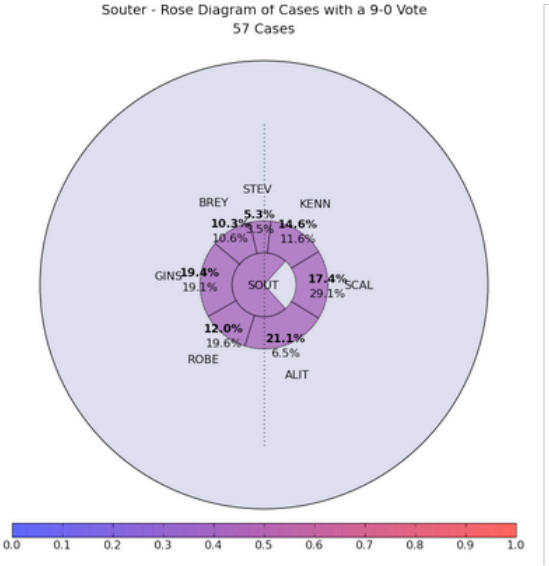

Scalia - Rose Diagram of Cases with a 9-0 Vote
57 Cases
BREY
SOUT
STEV
GINS
KENN
SCAL
ROBE
ALIT
9.9%
11.3%
9.3%
10.4%
9.8%
7.3%
22.8%
25.0%
14.4%
12.7%
14.6%
26.7%
19.2%
6.6%
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

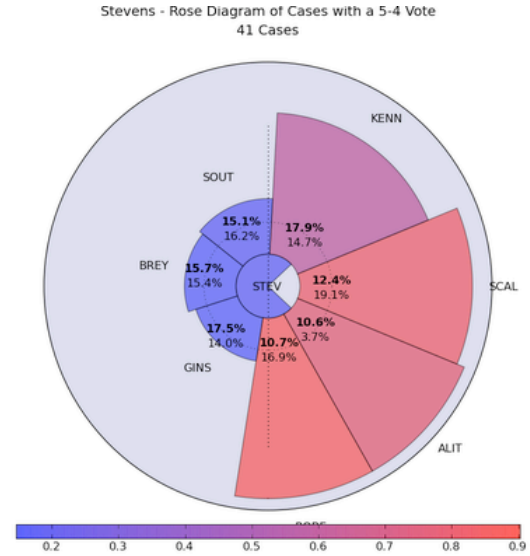

Souter - Rose Diagram of Cases with a 5-4 Vote
41 Cases
KENN
STEV
BREY
GINS
SCAL
ROBE
ALIT
SOUT
11.0%
9.4%
17.6%
15.7%
7.3%
7.9%
17.8%
29.8%
19.3%
16.8%
7.6%
13.1%
19.4%
7.3%
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

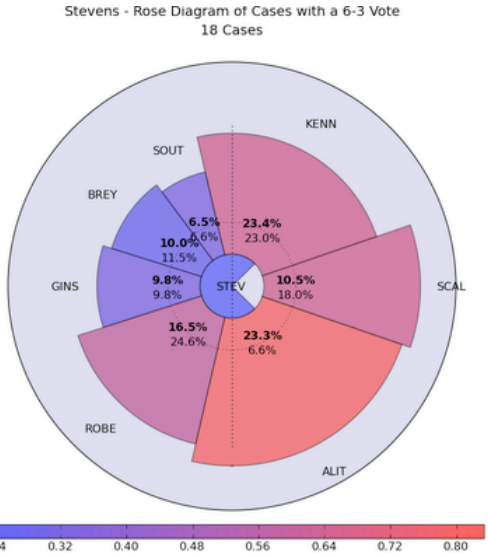

Souter - Rose Diagram of Cases with a 6-3 Vote
18 Cases
KENN
STEV
BREY
GINS
SCAL
ROBE
ALIT
SOUT
16.0%
9.8%
12.6%
11.0%
9.5%
9.8%
24.8%
37.8%
16.3%
14.6%
11.0%
14.6%
9.7%
2.4%
0.24 0.32 0.40 0.48 0.56 0.64 0.72 0.80

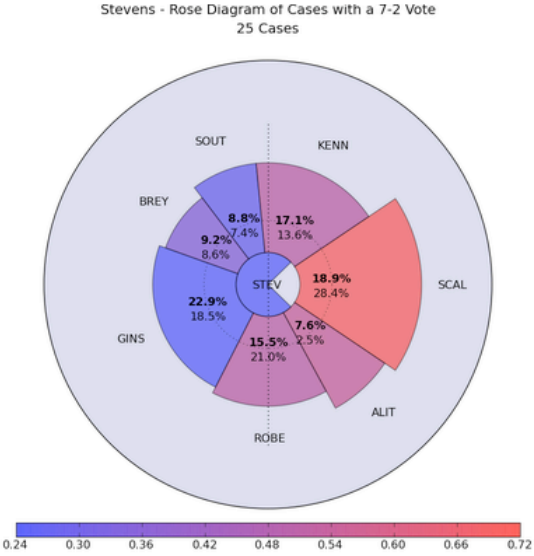

Souter - Rose Diagram of Cases with a 7-2 Vote
25 Cases
STEV
BREY
KENN
GINS
SCAL
ROBE
ALIT
SOUT
8.7%
6.5%
11.4%
12.1%
10.4%
9.3%
18.3%
16.8%
18.0%
30.8%
10.3%
15.9%
22.8%
8.4%
0.24 0.30 0.36 0.42 0.48 0.54 0.60 0.66 0.72

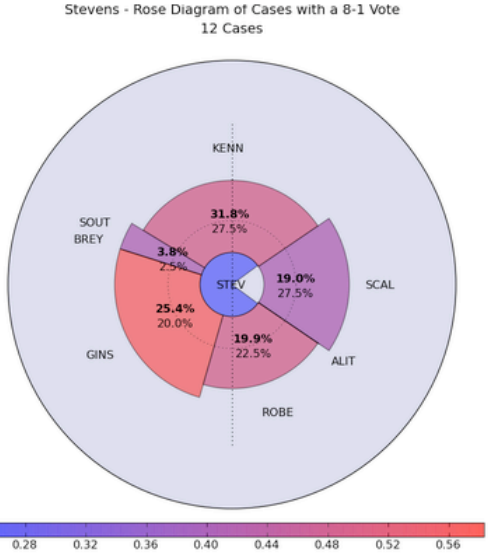

Souter - Rose Diagram of Cases with a 8-1 Vote
12 Cases
STEV
BREY
KENN
GINS
SCAL
ROBE
ALIT
SOUT
13.5%
10.3%
4.9%
7.7%
10.2%
10.3%
19.7%
17.9%
18.4%
30.8%
11.8%
15.4%
21.5%
7.7%
0.28 0.32 0.36 0.40 0.44 0.48 0.52 0.56

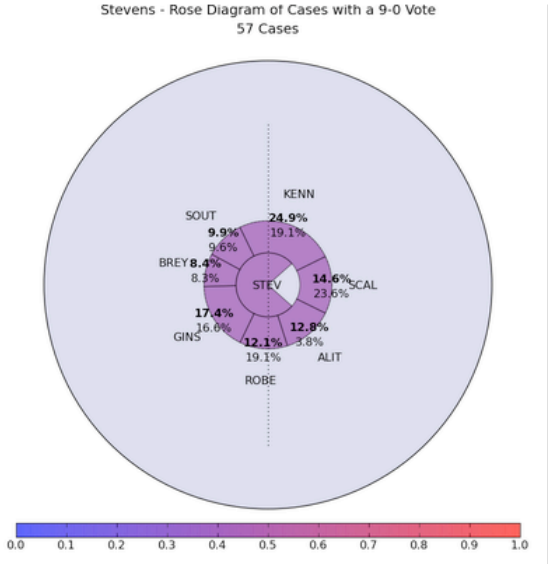

Souter - Rose Diagram of Cases with a 9-0 Vote
57 Cases
STEV
BREY
KENN
GINS
SCAL
ROBE
ALIT
SOUT
5.3%
3.5%
10.3%
10.6%
14.6%
11.6%
19.4%
19.1%
17.4%
29.1%
12.0%
19.6%
21.1%
6.5%
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0


Stevens - Rose Diagram of Cases with a 5-4 Vote
41 Cases
KENN
SOUT
BREY
GINS
SCAL
ALIT
STEV
15.1%
16.2%
17.9%
14.7%
15.7%
15.4%
12.4%
19.1%
17.5%
14.0%
10.7%
16.9%
10.6%
3.7%
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9


Stevens - Rose Diagram of Cases with a 6-3 Vote
18 Cases
KENN
SOUT
BREY
GINS
SCAL
ROBE
ALIT
STEV
6.5%
2.6%
23.4%
23.0%
10.0%
11.5%
9.8%
9.8%
10.5%
18.0%
16.5%
24.6%
23.3%
6.6%
0.24 0.32 0.40 0.48 0.56 0.64 0.72 0.80


Stevens - Rose Diagram of Cases with a 7-2 Vote
25 Cases
SOUT
KENN
BREY
GINS
SCAL
ALIT
ROBE
STEV
8.8%
7.4%
17.1%
13.6%
9.2%
8.6%
18.9%
28.4%
22.9%
18.5%
7.6%
2.5%
15.5%
21.0%
0.24 0.30 0.36 0.42 0.48 0.54 0.60 0.66 0.72


Stevens - Rose Diagram of Cases with a 8-1 Vote
12 Cases
KENN
SOUT
BREY
GINS
SCAL
ALIT
ROBE
STEV
31.8%
27.5%
3.8%
2.5%
19.0%
27.5%
25.4%
20.0%
19.9%
22.5%
0.28 0.32 0.36 0.40 0.44 0.48 0.52 0.56


Stevens - Rose Diagram of Cases with a 9-0 Vote
57 Cases
KENN
SOUT
BREY
GINS
SCAL
ALIT
ROBE
STEV
24.9%
19.1%
9.9%
9.6%
8.4%
8.3%
14.6%
23.6%
17.4%
16.0%
12.8%
3.8%
12.1%
19.1%
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Appendix B Discourse Markers

Note: Some of these discourse markers include some regular-expression syntax.

above all
above all
absolutely
accordingly
actually
add to this
additionally
admittedly
after
after all
after that
after this
afterwards
again
again and again
albeit
all in all
all right
all the same
all this time
already
alright
also
also because
alternatively
although
altogether
always
assuming that
analogously
and
and again
and also
and another
and then
another time
anyhow
anyway
apart from
apart from that
arguably
as
as a
consequence
as a corollary
as a hypothetical
as a logical
conclusion
as a matter of
fact
as a result
as a whole
as against
as an
as briefly as
as closely as
as evidence
as far as
as for
as i said
as i say
as i understand
as if
as it happened
as it is
as it turned out
as long as
as luck would
have it
as soon as
as such
as though
as to
as we shall
as we will
as well
aside from
assuming
at a time
at any rate
at first
at first sight
at first view
at last
at least
at most
at once
at some level
at some point
at that
at that moment
at that point
at that time
at the moment
at the moment
when
at the outset
at the same time
at the time
at this date
at this moment
at this point
at this stage
at which
at which point
back
back to my
original point
because
because of
because of this
before
before long
before that
before then
besides
besides that
better
briefly
but
but also
but then
but then again
by
by all means
by and by
by and large
by comparison
by contrast
by that time
by the same
by the same
token
by the time
by the way
by then
certainly
clearly
come to think of
it
conceivably
consequently
considering
considering that
contrariwise
conversely
correspondingly
decidedly
definitely
despite
despite that
despite the fact
that
despite this
doubtless
each time
earlier
either
either case
either event
either way
else
elsewhere
equally

especially
essentially
even
even after
even before
even if
even so
even then
even though
even when
eventually
ever since
every time
everywhere
evidently
except
except after
except before
except if
except when
except in so far as
except that
except when
excuse me
failing that
finally
fine
first
first of all
firstly
following
following this
for
for a start
for example
for fear that
for instance
for one
for that
for that matter
for that reason
for the reason that
for the simple reason
for this
for this reason
for me
formerly
fortunately
frankly
from all
from everything
from now on
from then on
from your
answer
further
furthermore
given
given that
granted that
having said
having said that
hence
here
herein
here's
heretofore
hitherto
however
however that
may be
hum
i don't think
i guess
i mean
i say
i suppose
i suspect
i take it
i think
i thought
i understand
if
if ever
if in fact
if indeed
if not
if only
if so
if such a
in a different vain
in a sense
in actual fact
in addition
in all candor
in all due respect
in any case
in any event
in case
in comparison
in conclusion
in consequence
in contrast
in doing
in doing so
in doing this
in effect
in essence
in fact
in fairness
in general
in just the same way
in may be concluded that
in my case
in my opinion
in my view
in one instance
in order to
in other respects
in other words
in our judgment
in our view
in part
in particular
in place of
in point of fact
in practice
in real world terms
in response
in retrospect
in short
in so doing
in so many words
in spite of
in spite of that
in such a
in such an
in sum
in that
in that case
in that instance
in that respect
in that scenario
in that statement
in the beginning
in the case of
in the end
in the event
in the first place
in the hope that
in the meantime
in the same way
in theory
in this case
in this connection
in this respect
in this way
in truth
in turn
in which
in which case
in your opinion
in your view
inasmuch as
incidentally
including
incontestably
incontroversialy
indeed
indisputably
indubitably
initially
insofar
insofar as

instantly
instead
instead of
interestingly
interestingly enough
ironically
it becomes
it can be concluded that
it follows
it follows that
it happens
it is because
it is clear
it is conceivable
it is conclusive
it is correct
it is for this reason
it is only
it (may|might) seem that
it (may|might) appear that
it (may|might) seem that
it turns out
just
just a pause
just about
just again
just as
just before
just then
kind of
largely
largely because
last
lastly
later
lest
let us
let us assume
let us consider
like
likewise
listen
literally
look
luckily
mainly
mainly because
meanwhile
merely
merely because
mind you
more accurately
more
importantly
more precisely
more
specifically
more to the
point
moreover
most likely
much as
much later
much sooner
my point
my position
my question
my response
my solution
my
understanding
naturally
needless
neither
neither is it the
case
never again
nevertheless
next
next moment
next time
no
no doubt
no matter
no sooner than
nonetheless
nor
normally
not
not at all
not
automatically
not because
not by itself
not completely
not directly
not exactly
not necessarily
not only
not quite
not really
not specifically
not that
notably
notwithstanding
notwithstanding
that
now
now that
obviously
of course
oh
okay|ok
on a different
note
on account of
on another
on balance
on condition
on condition
that
on its face
on its own
on one hand
on one side
on that
on that point
on that question
on that very
point
on the bases
on the basis
on the contrary
on the face of
on the grounds
on the grounds
that
on the one hand
on the other
on the other
hand
on the other side
on this basis
on this
particular issue
on top of it
on top of that
on top of this
on which
once
once again
once more
only
only after
only because
only before
only if
only when
oops
or
or again
or else
ordinarily
originally
other than
otherwise
our focus
our only point
our point
our position
overall
parenthetically
particularly
particularly
when
perhaps
plainly
possibly

potentially
practically
precisely
presently
presumably
presumably
because
previously
probably
provided
provided that
providing that
put another way
quite
quite likely
quite simply
quite the contrary
rather
reasonably
reciprocally
regardless
regardless of
that
returning to
right
rightly so
say
second
secondly
see
seeing as
seeing that
seemingly
significantly
similarly
simply
simply because
simultaneously
since
so
so far
so if
so that
some time
soon
speaking of
specifically
still
still and all
strictly speaking
subsequently
such as
such that
suddenly
summarizing
summing up
suppose
suppose that
supposedly
supposing that
sure enough
surely
technically
that
that done
that is
that is all
that is how
that is to say
that is why
that reminds me
that said
that way
the end
the fact is
the fact is that
the first time
the instant
the issue here
the key
the key words
the last time
the latter
the logic is that
the moment
the more
the more often
the next time
the one time
the point
the point being
the point is
the question
the question is
the thing is
then
then again
theoretically
there again
there are a few
things
thereafter
thereby
therefore
there('s| is) no doubt
thereupon
third
thirdly
this case
this claim
this court
this means
this time
though
thus
thus far
to add
to be clear
to be fair to
them
to be precise
to be sure
to begin with
to clarify
to close
to comment
to conclude
to explain
to follow-up
to get back
to go on
to go to
to illustrate
to interrupt
to make matters worse
to me
to my knowledge
to note
to open
to put it
to put it in context
to put it this way
to repeat
to start with
to stop
to sum up
to summarize
to take an example
to the best of my knowledge
to the best of our knowledge
to the degree that
to the extent
to the extent possible
to the extent that
to this end
to the assumption
too
traditionally
two
two answers
two points
two primary reasons
two reasons
two responses
two separate
two things
typically
uh
ultimately
undeniably
under the circumstances

under these circumstances
understand
undoubtedly
unfortunately
unless
unquestionably
until
until then
up to now
up to this
very briefly
very likely
very quickly
we agree
we believe
we believed
we might say
we think not
we think that
well
what i mean to say
what is more
whatever
when
whenever
where
whereas
whereby
whereupon
wherever
whether
whether or not
which
which is why
which means
which reminds me
whichever
while
while i have you
who
whoever
with absolute certainty
with all due respect
with all respect
with one addition
with regard to
with respect
with respect to
with that
with this
without
yes
yet
you know
you see
false
true

References

Ali v. Federal Bureau of Prisons. 06-9130 U. S. (2007).

Benesh, S. C. (2002). Becoming an Intelligent User of the Spaeth Supreme Court Databases. *Southwestern Political Science Association Meeting*. New Orleans, LA.

Biscupic, J. 2006. Justices make points by questioning lawyers. *USA Today*. (Oct. 5, 2006).

Brown, G. and Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.

Clopper, C. J., and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26, 404–413.

Cortes, C. and Vapnik, V. (1995). Support-vector network. *Machine Learning*, 20.

Draper, G. M., Livnat, Y., Riesenfeld, R. F. (2009). A Survey of Radial Methods for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics.* 15(5), 759-776.

Duke Law. (2009). Supreme Court Associate Justice Antonin Scalia presides over Dean's Cup Moot Court Competition Duke Law News and Events. http://www.law.duke.edu/news/story?id=2943&u=11.

Engquist v. Oregon Dept. of Agriculture. 07-474 U. S. (2008).

Evans, M., McIntosh, W., Lin, J., and Cates, C. (2007). Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research. *Journal of Empirical Legal Studies*, 4(4), 1007-1039.

Federal Election Comm'n v. Wisconsin Right to Life, Inc.. 06-969 U. S. (2007).

Forbes-Riley, K. and Litman, D. (2004). Predicting Emotion in Spoken Dialogue from Multiple Knowledge Sources. In *Proceedings of the Human Language Technology Conference: 4th Meeting of the North American Chapter of the Association for Computational Linguistics*.

Galley M., McKeown, K., Hirschberg, J., Shriberg, E. (2004). Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics* (669-676).

Garside, R. (1987). The CLAWS Word-tagging System. In: R. Garside, G. Leech and G. Sampson (eds), *The Computational Analysis of English: A Corpus-based Approach*. London: Longman.

Greene, S. and Resnik, P. (2009). More Than Words: Syntactic Packaging and Implicit Sentiment. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Grosz, B. and Hirschberg, J. (1992) Some Intonational Characteristics Of Discourse Structure. In *Proceedings of the International Conference on Spoken Language Processing*.

Grosz, B. and Sidner, C. L. (1986). Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3), 175-204.

Gurevych, I., Strube, M. (2004) Semantic Similarity Applied To Spoken Dialogue Summarization. In *Proceedings of the 20th International Conference on Computational Linguistics*.

Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hawes. T., Lin J., and Resnik, P. (2009) Elements of a Computational Model for Multi-Party Discourse: The Turn-Taking Behavior of Supreme Court Justices. *Journal of the American Society for Information Science and Technology*, 60(8), 1607 – 1615.

Hutchby, I. and Wooffitt, R. (2008). *Conversation Analysis*. Cambridge: Polity Press.

Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A. and Wooters, C. (2003). The ICSI Meeting Corpus. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (364–367).

Johnson, T. R. (2001). Information, oral arguments, and Supreme Court decision making. *American Politics Research*, 29(4), 331–351.

Johnson, T. R. (2004*). Oral arguments and decision making on the United States Supreme Court*. Albany, NY: State University of New York Press.

Johnson, T. R., Black, R., Goldman, J. and Treul, S. (2009) Inquiring Minds Want to Know: Do Justices Tip Their Hands with Questions at Oral Argument in the U.S. Supreme Court?. *Washington University Journal of Law & Policy*, 29.

Johnson, T. R., Black, R., and Ringsmuth, E. (2009) Hear Me Roar: What Provokes Supreme Court Justices to Dissent from the Bench? *Minnesota Law Review*.

Johnson, T. R., Spriggs, J. F., and Wahlbeck, P. J. (2007). Supreme Court Oral Advocacy: Does it affect the Justices' Decisions?. *Washington University Law Review*, 85.

Johnson, T. R., Wahlbeck P.J., and Spriggs, J.F., II. (2006). The influence of oral arguments on the U.S. Supreme Court. *American Political Science Review*, 100(1), 99–113.

Johnson, T. R., Wahlbeck, P. J., and Spriggs, J. F. (2006). The Influence of Oral Arguments on the U.S. Supreme Court, *American Political Science Review*.

Johnstone, B. (2007). *Discourse Analysis*. Malden: Blackwell Publishing.

Jovanovic, N., and Akker, R. op den. (2004). Towards automatic addressee identification in multi-party dialogues. In M. Strube and C. Sidner (Eds.), *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT/NAACL 2004* (89–92).

Kansas v. Marsh (Reargued). 04-1170 U. S. (2006).

Kohavi, R. (1995). The Power of Decision Tables. In *8th European Conference on Machine Learning* (174-189).

Kurland, P. B., & Hutchinson, D. J. (1983). The business of the Supreme Court, O. T. 1982. *The University of Chicago Law Review*, 50(2), 628-651.

Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In C.E. Brodley and A.P. Danyluk (Eds*.), Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)* (282–289).

Laver, M., Benoit, K., and Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2), 311–331.

MacWhinney, B., Bird, S., Cieri, C., and Martell, C. (2004). TalkBank: Building an open unified multimodal database of communicative interaction. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*.

Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

Marcu, D. (1997). The rhetorical parsing of unrestricted natural language texts. In P.R. Cohen and W. Wahlster (Eds.), *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL 1997)* (96-103), Madrid, Spain: ACL.

Marcu, D. and Echihabi, A. (2002) An Unsupervised Approach to Recognizing Discourse Relations. In *Proceeding of the ACL/NAACL*.

Martin, A. D. and Quinn, K. M. (2002). Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis*. 10, 134-153.

Michael A. Watson v. United States. 06-571 U. S. (2007).

Morris, J. and Hirst, G. (1991). Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text . *Computational Linguistics* 17(1), 21-48.

Mosteller, F. and Wallace, D. L. 1964. *Inference and Disputed Authorship: The Federalist*. Reading: Addison-Wesley.

Oates, S. (2001). A listing of discourse markers. *Technical Report ITRI-01-26*. Retrieved January 10, 2008, from University of Brighton, Information Technology Research Institute Web site: ftp://ftp.itri.bton.ac.uk/reports/ITRI-01-26.pdf.

Pang, B. and Lee, L. (2008). *Opinion Mining and Sentiment Analysis*. Boston: Now Publishers Inc.

Purver, M., Körding, K., Griffiths, T. and Tenenbaum, J. (2006). Unsupervised Topic Modeling for Multi-Party Spoken Discourse. In *Proceedings of COLING/ACL 2006* (pp. 17-24), Sydney, Australia: July 2006.

Randall v. Sorrell. 04-1528. U. S. (2004).

Rehnquist, W.H. (2002). *The Supreme Court*. New York: Vintage.

Rohde, D. and Spaeth, H. (1976). *Supreme Court Decision Making*. San Francisco: Freeman.

Rombeck, T. (2002). Justice takes time for Q&A. *Lawrence Journal-World*.

Ruger, T. W., Kim, P., Martin, A. D. and Quinn, K. M. (2002). The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking. *Columbia Law Review*.

Ruger, T. W., Kim, P., Martin, A. D. and Quinn, K. M. (2004). Competing Approaches to Predicting Supreme Court Decision Making. *Perspectives on Politics Symposium*. 2(4).

Samson v. California. 04-9728 U. S. (2006).

Schegloff, E. A. (2007). *Sequence Organization in Interaction: Volume 1: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Schiffrin, D., Tannen, D. and Hamilton, H. E. (eds.) 2001. *The Handbook of Discourse Analysis*. Malden: Blackwell Publishers Inc.

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

Settles, B. (2004). Biomedical named entity recognition using conditional random fields and rich feature sets. In N. Collier, P. Ruch, and A. Nazarenko (Eds.), In *Proceedings of the COLING 2004 International Joint workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP 2004)* (107–110).

Sha, F., and Pereira, F. (2003). Shallow parsing with conditional random fields. In M. Hearst and M. Ostendorf (Eds.), In *Proceedings of Author Proof the 2003 Human Language Technology Conference and the North American Chapter of the Association for Computational Linguistics Annual Meeting* (134–141), Edmonton, Alberta, Canada: ACL.

Shullman, S. L. (2004). The illusion of devil's advocacy: How the justices of the Supreme Court foreshadow their decisions during oral argument. *The Journal of Appellate Practice and Process*, 6, 271–293.

Small v. United States. 03-750 U. S. (2004).

Snyder v. Louisiana. 06-10119 U. S. (2007).

Spaeth, H. J. (2009). The Original U.S. Supreme Court Judicial Database. http://www.cas.sc.edu/poli/juri/sctdata.htm.

Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Van Ess-Dykema, C. , Ries, K., Shriberg, E., Jurafsky, D., Martin, R. and Meteer, M. (2000). Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computatuinal Linguistics*, 26(3).

Sutton, C. and McCallum, A. (2006). Introduction to Conditional Random Fields for Relational Learning In L. Getoor and B. Taskar (Eds.), *Introduction to Statistical Relational Learning.*

Thomas, M., Pang, B., and Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In D. Jurafsky and E. Gaussier (Eds.), *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)* (327–335). Sydney, Australia: ACL.

Toutanova, K., Klein, D., Manning, C., and Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL* 2003 (252-259).

Travelers Casualty & Surety Co. of America v. Pacific Gas & Elec. Co. 05-1429 U. S. (2007) .

Wrightsman, L. S. (2008). *Oral Arguments Before the Supreme Court*. New York: Oxford University Press.

Wrightsman, L. S. (2008). *Oral Arguments Before the Supreme Court*. Oxford: Oxford University Press.

Yuan, J. and Liberman, Mark. (2008). Speaker Identification in the SCOTUS corpus. In *Proceedings of Acoustics '08*.