Modelling Intergroup Behaviour in Folketinget using NLP

Morten Ravnsgaard Gade (201706692)
Gustav Huitfeldt Kragelund Helms (201910185)

Supervisor: Riccardo Fusarolli

School of Communication and Culture, Aarhus University
Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark

All data and code for the analysis can be found at GitHub

January 3, 2022

Abstract

(Morten) Optimal Distinctiveness Theory states that individuals satisfy their two opposing needs for inclusion and distinctiveness through membership of optimally distinct groups. Empirical support for the theory has mostly consisted of correlations between phenomenological rapports and observed in-group bias in experimental settings. In this paper, we will take a different approach, where we will attempt to use linguistic behaviour as a measure of intergroup dynamics. Due to the availability of natural language produced in Folketinget, we focus on this specific social arena, and investigate if political parties in Folketinget behave according to predictions made from Optimal Distinctiveness Theory. We found partial correspondence between measurements and predictions, as previous levels of distinctiveness showed a credible effect on current levels of inclusion. The results suggest groups in Folketinget adjust levels of inclusion as opposed to levels of distinctiveness to maintain equilibrium. However, measuring group dynamics through NLP methods is novel and more future research should be conducted to further validate the findings.

Keywords— Optimal distinctiveness theory, Wordshoal, NLP, intergroup behaviour, social cognition

1 Introduction

(Gustav) A common conception of contemporary western society is that its members are becoming increasingly individualistic and seek to free themselves from groups they are dependent on (Dubois and Beauvois, 2005; Leonardelli et al., 2010). They strive to pursue their personal goals and advance their careers in a world that is centred around the individual. Meanwhile, it is widely accepted that the survival of the human species has been much dependent on cooperative group living, in which benefits like shared resources, division of labour and mutual protection could be exploited (Brewer, 1997; Leonardelli et al., 2010). Cooperative living is typically ascribed as a key trait to the sociality of humans, but how can this coexist with the individualistic tendencies of western societies? In the next sections, the most prominent theories on social cognition and intergroup behaviour will be reviewed to illuminate how the two opposing needs of individualism and belonging coexist in groups and what influence they have on the behaviour of the group. Subsequently, we investigate possible methods for measuring intergroup behaviour.

1.1 Key concepts of intergroup behaviour

(Morten) In order to investigate group dynamics we will first introduce common concepts within the domain. Tajfel et al., 1979 redefined much of the terminology used within the domain with their proposal of *Social Identity Theory*. They define *social groups* in a non-restrictive sense, mentioning two sufficient conditions for group membership. Firstly, the individual has to internalise the membership. In other words, it is not sufficient that the individual is seen as a group member only by other individuals. Secondly, other group members likewise have to see the individual as part of the group. *Intergroup behaviour* is displayed when the identification with

an *in-group* drives, or determines, the individual's behaviour towards a collection of individuals perceived as an *out-group*. Simply put, the in-group is the group the actors identify themselves with, and the out-group is the external collection of individuals acted towards. Another important concept is *in-group bias*, which is the favouritism displayed towards one's own group. The degree of displayed favouritism is often interpreted as the strength of the tie between the individual and the in-group (Brewer, 1979).

Behaviour can be positioned on a continuum ranging from *interpersonal* to *intergroup* behaviour. Tajfel et al., 1979, provides examples of behaviour for each extreme. Pure interpersonal behaviour is only driven by individual traits and characteristics of the interacting agents, which is similar to an interaction between husband and wife. In contrast, pure intergroup behaviour is only driven by the identification with an in-group. Extreme intergroup behaviour is often exemplified with how soldiers from different countries act towards each other. Another continuum, important for social cognition, is that of social mobility vs. social change. This is a continuum of how the individual perceives the intergroup structure. If an individual perceives the societal structure of the groups as one corresponding to the 'social mobility' end of the continuum, they believe groups are permeable. If the society of groups is perceived as 'social change', they are stratified, and boundaries between groups are in a sense objective. According to Tajfel et al., 1979, the perception of a culture subscribing to 'social change' will correlate positively with the intensity of intergroup behaviour.

1.2 Social Identity Theory

(Gustav) Fundamental to his theory on group behaviour, Tajfel et al., 1979, introduced the concept of *social identity*. Social identity is defined as an individual's perceived identification with a group and the connotations and values of the group (Tajfel et al., 1979). Thus, an individual can have either positive or negative social identity. If positive, the individual ascribes positive values to a specific membership of a group and if negative the opposite. Individuals consistently strive towards a positive social identity, which is affected through comparison with relevant out-groups. When the result of such comparison is negative, the individual's self-esteem suffers. A negative social identity would motivate the individual to seek *positive distinctiveness* on some relevant, value-laden dimension, and the perceived importance of the dimension affects the degree to which the comparison changes the individual's self-esteem. For example, universities do not weigh comparisons on food quality with other universities as much as comparisons regarding average GPA. Furthermore, only comparisons made with relevant out-groups affect social identity. The university does not compare itself with kindergartens and certainly not political parties. In summary, individuals seek positive distinctiveness, measured in the comparison between in-group and relevant out-groups on value-laden dimensions. This strive to obtain positive distinctiveness elicits behaviour in situations where the membership of the individual is internalised and the relevancy, or comparability, of the out-groups has to be high.

(Morten) Previous research on social identity theory has provided empirical support for how threats to an individual's perceived intergroup distinctiveness will trigger a reaction, which attempts to restore the positive distinctiveness (Jetten et al., 2004). In such a case, a number of strategies to restore positive distinctiveness has been theorised. Tajfel et al., 1979, offers four strategies. Firstly, the dissatisfied individual can seek other groups with higher perceived status. This assumes the individual perceives a group culture which is positioned in the 'social mobility' end of the continuum, in which the boundaries between the groups are permeable. If the culture is perceived as stratified with objective, clear-cut group boundaries, it significantly reduces the efficacy of this strategy. Importantly, this strategy is egotistical in the sense that it may lead to a disintegrative process, in which the group dissolves. Secondly, the group may attempt to redefine value-laden dimensions, by assigning positive values to before negative values on a relevant dimension of comparison. For instance, an underprivileged group could assign a positive value to being underprivileged and thereby maintain positive distinctiveness in being underprivileged. The third strategy is that groups can simply change the out-group against which comparisons are made. Lastly, groups might seek to compete with out-group(s) and seek positive distinctiveness on the relevant dimensions of comparison through competition. Generally, the strategies can be split into two categories. The first being the strategies in which individuals are seeking positive distinctiveness through a change in group membership. The second, is the strategies in which individuals seek to improve their positive distinctiveness through their current group membership.

1.3 Optimal Distinctiveness Theory

(Gustav) Social identity theory predicts that groups primarily seek distinction from other groups in comparisons with these. Social identity theory is thereby only considering intergroup relations as motivational for group identification. The intragroup effects of intergroup behaviour is seen as a resulting consequence by Tajfel et al., 1979, as they write intergroup competition enhances intragroup cohesiveness. This model does not allocate serious thought to the intragroup effects, and the theory of group identification does not detail the motivational factors underlying the need for a social identity. It is not entirely clear why group members strive

for distinctiveness, despite the empirical support for the theory. The obvious explanation is human nature, i.e. some inherent drive to *be someone* in relation to others. Assuming only this drive, it seems equally obvious to conclude that we seek to differentiate ourselves from everyone else, why group membership suddenly seems paradoxical.

(Morten) In a paper titled "The social self: On being the same and different at the same time", Brewer, 1991, raises this concern, stating that social identity is essentially a more inclusive self-definition compared to the self in the most restrictive sense. Some social identities are more inclusive than others. John Doe as an American citizen is a more inclusive social identity vs. John Doe as a university student at Stanford University. In the former case, John Doe is deindividuated to a higher degree compared to the latter. However, the sense of inclusiveness in an individual's social identity is context dependent. Different contexts can contract or expand the sense of inclusiveness. For example when studying abroad, one's nationality is a bigger part of one's social identity. Moving through different levels of inclusion affects the frame of reference. John as an American citizen will compare himself to other countries on dimensions such as welfare, economic prosperity and such, whereas John the student compares himself and his fellow students to students from other universities.

(Gustav) As a consequence of the inadequacy of describing social identity only in terms of intergroup comparisons Brewer proposes *Optimal Distinctiveness Theory* (ODT). Much research finds the need for inclusion and the need for distinctiveness compete with each other. Individuals accept an intermediate level of both, such that both needs are met to a satisfactory degree (Snyder and Fromkin, 1977, 2012). ODT suggests these needs are satisfied on a group level, through membership in moderately inclusive, or optimally distinct, groups. The level of inclusion is determined through in-group relations, whereas distinctiveness is determined through comparisons between in-group and out-groups. Conceptually, this can be seen as two levels of comparison, each pertaining to a different need. Through group membership, we can be seen as similar on one level, and different on the other, hence being the same and different simultaneously. For instance, a social identity too inclusive triggers the need for uniqueness and vice versa. Individuals are therefore only at rest when identification provides both inclusion and distinctiveness. An interesting conclusion, which follows from these assumptions, is that one's personal identity, which represents individualization to the highest degree, does not meet the need for inclusion. Empirical support for this claim involves the concept of *fraternal deprivation*, where the in-group is perceived as deprived compared to an out-group. Fraternal deprivation seemed more motivational than *personal deprivation*, i.e. group identity takes precedence over personal identity (Dubé and Guimond, 1986).

(Morten) Leonardelli et al., 2010, emphasises how individuals are motivated to maintain an *equilibrium* between the two opposing forces. Discrepancies from the equilibrium is then solved through adopting new social identities closer to equilibrium. It is important to note the motivational force is still related to self-esteem, as social identity is the basis for self-evaluation. The authors highlight optimal distinctiveness is not to be seen as some property the group either has or not. Instead, optimality is achieved when the equilibrium is maintained on an individual level, albeit achieved through group membership of the individual and the individual's perceptions of this. This equilibrium is depicted in figure 1. In figure 1 the two lines represent the two opposing needs, where assimilation is the need for inclusion and differentiation is the need for distinctiveness. The x-axis represents the actual level of inclusion at the left figure and the level of distinctiveness in the right figure. The y-axis represents the level of the need for either inclusion or distinctiveness. Importantly, both the intercept and slope of all the lines may vary from context to context and between individuals. Considering the mentioned ideas, the level of inclusion within a group and the level of distinctiveness between groups will be positively correlated in optimally distinct groups. E.g. when a group has high inclusion within, it will excite the need for distinctiveness between groups and vice versa. Therefore, if both needs are satisfied, as in optimal distinct groups, the levels of inclusion and distinctiveness will be equally high. On a similar note, imbalances might be resolved through a multitude of strategies. Lack of inclusion can be resolved through increased deindividuation, i.e. more intragroup assimilation, or through increase of group size. Lack of distinctiveness might be achieved through less intragroup assimilation, lowering group-size or using one of the previously presented strategies for restoring positive distinctiveness.

(Gustav) The following sections seeks to elucidate how group behaviour and social cognition has usually been measured and what some of the underlying assumptions are. Firstly, the trade off between controlled laboratory experiments and ecological validity is highlighted. Secondly, the assumption about how group behaviour is measured at the level of its constituent parts and what the measures typically consist of. Lastly, zooming in on research on ODT and the measures typically used to study the theory.

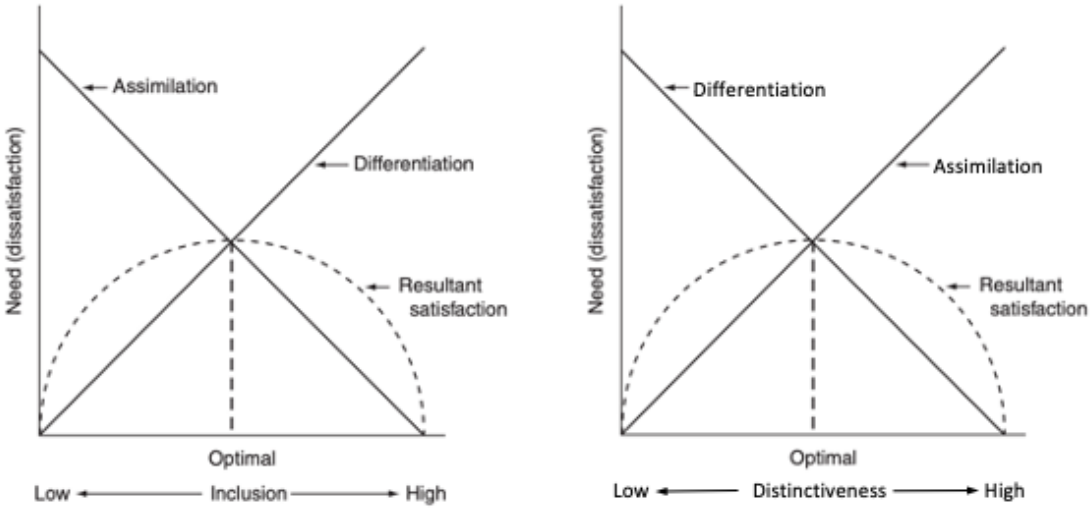

Figure 1: Optimal Distinctiveness Theory. The figure on the left shows the need for assimilation and differentiation at high and low levels of inclusion within the group. The figure at right shows the same for high and low levels of differentiation between groups (Brewer, 1979).

1.4 Ecological validity

(Morten) Measuring group behaviour is a discipline that involves a lot of methodological considerations to ensure that ecological validity is maintained outside of the laboratory. Ecological validity is typically referred to as whether or not observed behaviour in the laboratory generalises to natural behaviour found outside of the laboratory (Araujo et al., 2007; Bem and Lord, 1979; Schmuckler, 2001). Ecological validity is usually achieved by making a study design in which salient elements like stimulus or the task represent the natural world as much as possible. Meanwhile, studies seek to isolate variables of interest while simultaneously controlling for other influential variables to investigate a specific targeted behaviour. Ecological validity and control of variables is often a trade off relationship in which increasing one of the two often comes with the cost of the other. I.e. the more a behaviour is isolated in the laboratory the less it often resembles behaviour outside of the lab. Within the study of social identity, Brewer, 1991, argues that findings from the laboratory concerning social identity theory and ODT, should be tested in naturalistic settings in order to check the ecological validity of the findings. Testing previous findings of social identity and ODT in a naturalistic setting could further validate or invalidate the theory. This is among the reasons why the paper at hand seeks to investigate ODT in a naturalistic setting outside of the lab.

(Gustav) An assumption often made in research on intergroup behaviour is that the behaviour of the group is the sum of the behaviour of its constituent parts i.e. the members of the group. Because of this, much research on group behaviour has measured individual behaviour and used the summarised behaviour of the individuals within a group to extrapolate on the behaviour of the group. An example of a study on intergroup behaviour that utilises this assumption is a study conducted by Goff et al., 2008. In the study, participants were asked to sit down and have a 'comfortable conversation' with other in-group vs out-group members. The distance at which they placed their chair from the conversation partner were measured and used to extrapolate on intergroup behaviour. In this paper, we will also summarise group behaviour by measuring the behaviour of its constituent parts.

1.5 Types of measures

(Gustav) Previous studies on intergroup behaviour have mainly been conducted in laboratories in which there have usually been used three kinds of measures on individual behaviour: (1) rating measures, (2) behavioural measures and (3) allocation tasks (Böhm et al., 2020).

> *Rating measures* consist of questionnaires or self-reports on phenomenological experiences of group structures. An example of a study that uses rating measures is performed by Leondardelli and

Pickett, 2009, in which undergraduates were asked to rate the degree to which students within their university were similar and the degree to which other universities were similar to theirs. The reports were used as measures of the group structure.

Behavioural measures are measures of rather spontaneous intergroup behaviour on an individual level. An example of a study that uses behavioural measures could be the study conducted by Goff et al., 2008, which were highlighted in the previous section. Here, the distance with which participants placed their chairs from a conversational partner was the behavioural measure of a spontaneous behaviour.

Allocation measures are a kind of behavioural measure. However, they differ from behavioural measures by measuring decisions made on allocation of valuable units (e.g. money) between in-group and out-group members. Thereby, allocation measures are typically measuring behaviour in decision making situations that are forced upon the participant rather than the spontaneous behaviour described in the previous section (Böhm et al., 2020).

(Morten) As presented in earlier sections, a lot of the research on social identity and ODT has been conducted using rating measures. In these studies participants have reported their perceived experience of the structure within their group and the relationship between groups. The study conducted by Leondardelli and Pickett, 2009, is an example of this. These phenomenological reports have been used as measures of inclusion and distinctiveness and used to test ODT. As the aim of this paper is to investigate ODT in a naturalistic setting, we will use behavioural measures of inclusion and distinctiveness to ensure spontaneous and natural intergroup behaviour. By using behavioural measures rather than self-reports on subjects' perceived experience of group infrastructure, we avoid some of the main caveats related to phenomenological reports. Without going too much into the subject, some of the challenges typically relate to validation of the accuracy of a report or the risk of subjects being biassed (Lutz and Thompson, 2003; Olivares et al., 2015). See Smith, 2018, for a more detailed review. However, using behavioural measures instead of rating measures, comes with its own challenges. Behavioural measures often implies that the motivation underlying a measured behaviour is not uncovered (Böhm et al., 2020).

1.6 Formalisation of distinctiveness and inclusion

(Gustav) In experimental studies on ODT, distinctiveness has typically been formalised as the perceived difference between groups on a relevant dimension of comparison. However, there are some inconsistencies in terms of how inclusion is translated from a theoretical concept into a measurable variable. Previous studies have typically formalised inclusion in three different ways: as a measure of deindividuation, a measure of group size or a measure of assimilation within the group.

Brewer, 1991, designed an experimental paradigm, in which perceived inclusion was formalised as *deindividuation*. In the deindividuation condition, participants were told they were just generic members of a group. They should therefore behave as group members, not as individuals. In the control condition, such instructions were not given. Individuals were either assigned to the majority group or the minority group. Results showed strong in-group bias for individuals in the minority group when they've been deindividuated, but positive in-group bias for both groups in the control condition. Leonardelli et al., 2010 refer to multiple studies, varying in the interpretation of inclusion. In one paradigm, participants either belonged to a majority or minority, and their in-group bias was then measured. This paradigm adhered to a *group-size* definition of inclusion, in which the majority-group represented high inclusion and the minority-group represented low inclusion (McGuire and McGuire, 1988). In another paradigm, participants rate the similarity between themselves and in-group members, and this rating is implemented as the level of inclusion, meaning inclusion becomes a measure of perceived *assimilation* within the group (Leonardelli and Pickett, 2009).

(Morten) Many translations seem applicable. As a theory seeking to elucidate the motivational forces behind group identification, perceived deindividuation seems appropriate, since it best corresponds with Brewer's original definition. However, this measure relies on a phenomenological approach, with all the caveats of such an approach outlined previously. Using the group-size formalisation of inclusion is a more direct and accessible measure. However, it does not correspond very well with Brewer's definition of inclusion as perceived deindividuation. The assimilation formalisation does seem appropriate as the more assimilated members of a group become, the more their individual characteristics vanish. They become βtriggering the need for distinctiveness.

1.7 Folketinget as arena for intergroup behavior

(Gustav) In this study, the Danish parliament *Folketinget* has been chosen as the arena to investigate intergroup behaviour. Folketinget has several traits that makes it attractive to do intergroup research on. Firstly,

Folketinget is very well documented. All debates in Folketinget have been transcribed since 1953 and are publically available. Secondly, groups within Folketinget are clearly defined, since almost all members of Folketinget are associated with a political party. Furthermore, the group structure of Folketinget has features that are very compelling when doing research on intergroup behaviour. All groups are political organisations, which indicates that any intergroup comparisons made between members are most likely going to be on a relevant dimension (e.g. political beliefs). This is an important feature as it is a premise for intergroup behaviour motivated by social identity, cf. social identity theory by Tajfel et al., 1979. Moreover, groups in Folketinget are very robust. They rarely dissolve and members do not often change parties. This means that the environment of the groups in Folketinget would be categorised as a *social change* group environment. To summarise, social change is an environment in which groups are stratified and according to Tajfel et al., 1979 a social change group environment intensifies intergroup behaviour. The stratification in Folketinget is partly due to the fact that political parties are often characterised by a political belief or ideology. Members of parties are typically motivated to join a party by a correspondence between the political belief of the member and the party. The political beliefs of parties are often contradicting why a change of party is limited, as there is only a limited supply of parties that represent the political beliefs of the individual. An environment in which intergroup behaviour is amplified is a good environment for studying intergroup behaviour as it increases chances of measuring subtle effects. Another trait that makes Folketinget attractive is that groups in Folketinget are all operating on the same task that is to process bills (i.e. legislative proposals). Also, the practical procedure of bills has well defined rules. The common task and its procedural rules ensures that the behaviour from members within Folketinget are comparable as we assume the behaviour concerns the same task and follows some common rules. This should control for a lot of unsystematic variance.

1.8 Hypotheses

(Morten) If the groups within Folketinget behave according to the previously presented theory on ODT, we would expect to find that intergroup distinctiveness is positively associated with intragroup inclusion (Leonardelli et al., 2010). In other words, if the inclusion within a party increases, we would expect it to be followed by an increase in distinctiveness between parties. Likewise, if the distinctiveness between a party and other parties increases, we would expect it to be followed by a rise in the level of inclusion within a party. The scope of this paper is to examine whether intergroup behaviour of political parties within Folketinget acts according to predictions made from ODT. As ODT would predict a positive association between inclusion and distinctiveness we hypothesise that:

> **H1:** The current level of inclusion within a party is positively associated with previous levels of distinctiveness between that party and other parties.
>
> **H2:** The current level of distinctiveness between a party and other parties is positively associated with previous levels of inclusion within the party.

In the next sections, different methods on how to measure positions from textual data will be presented and reviewed.

1.9 Measuring positions from textual data

(Gustav) In Goet, 2019, prominent methods used to measure preference variation from textual data in the context of parliamentary speeches are reviewed. Recent published literature on the topic has used either unsupervised models and supervised models. As reviewed by Goet, the most prominent unsupervised model is the *Wordshoal* method which was first presented by B. E. Lauderdale and Herzog, 2016. The most prominent supervised model is a machine classifier presented by Peterson and Spirling, 2018. The two methods were conducted and tested on a common validation framework designed by Goet, 2019, in which the machine classifier performed best in the context of polarisation measurement. Wordshoal, however, gives more nuanced estimates of preference variation which allows for more elaborate analysis. Even though it traditionally has been used as a tool for measurement of preference variation, the estimates generated through the method can be applied in other contexts. In this paper we apply the Wordshoal method because it enables analysis of the intergroup dynamics in Folketinget. In the discussion section, we use the machine classifier to validate the measures obtained through the Wordshoal method, as the machine classifier performed superiorly in Goet's validation framework.

(Morten) Wordshoal as a method is an extension of Wordfish (Slapin and Proksch, 2008). Wordfish is a statistical model that can estimate a latent position of a speaker based on word frequencies. This means that Wordfish can distribute speakers on a one-dimensional scale by comparing the overlap in word use between speakers. The challenge with Wordfish is to ensure that the measured variance between the word use of speakers correspond to the actual variance it intends to measure. There are several sources of variation in word use.

B. E. Lauderdale and Herzog, 2016 has tried to arrange the most prominent sources of variation in roughly descending order by influence. These are: (1) Language, (2) Dialect, (3) Topic and only then (4) Position. This means that to use Wordfish to measure position one has to account for higher sources of variation in the hierarchy. The contribution of Lauderdale Herzog propose a way of accounting for variation caused by topic. They make the assumption that speeches from the same day will regard the same topic. The idea is that if you constrict a Wordfish scaling to a single day then you would have accounted for variation in word use caused by topic and the resulting estimates will then reflect the positions of the speakers rather than variation caused by different topics. They call this method *Wordshoal* and it is a two-step process. First step consists of generating Wordfish estimates for speakers per day. Second step is to use a Bayesian factor analysis to construct a common scale for the estimates generated in the first step. Since there will be a lot of speakers who do not speak in all debates there will also be a lot of Wordfish estimates missing, which means that the simplest factor analysis methods do not apply. A fully Bayesian treatment of the factor model will help overcome the problem with missing observations, by treating all missing observations as missing at random (B. E. Lauderdale and Herzog, 2016). It is thereby possible to estimate the position of all speakers in a corpus' position on a one dimensional latent scale with Wordshoal where topic variation is accounted for.

In summary, we've chosen Folketinget as a social arena in which we seek to investigate intergroup behaviour. This paper will implement a non-traditional approach, as we will use behavioural measures obtained through the Wordshoal method. Next, we will outline the process of data collection and give a more detailed description of how the Wordshoal method was implemented and the estimates modelled.

2 Methods

2.1 Data inquiry

(Gustav) Folketingstidende.dk has made transcripts of parliamentary debates publicly accessible through their website on which you can access transcripts from Folketinget from 1953 till date. The quality of the transcripts vary across time, with the oldest being of the worst quality. We've chosen to work with transcripts in the period 1990 - 2021, as pdfs from before this period are formatted very differently. In the selected time period transcripts are structured similarly, enabling us to process each file in similar fashion. These transcripts are available in pdf format, which means that the conversion from human-readable format to machine-readable format includes multiple steps. The conversion process used in this paper is illustrated in figure 2.

In total, 3270 pdfs were scraped and segmented into 810.736 speeches. The process of gathering the data can be summarised as follows: firstly, all files were downloaded. Secondly, text was extracted from each pdf. Thirdly, the text was segmented into speeches. Lastly, the speeches were preprocessed and combined with metadata. We revised and adopted scripts originally written by Malte Lau Petersen for the Danish Gigaword Project (Strømberg-Derczynski et al., 2020). Next, the process of gathering and preparing the data for analysis are presented in detail.

The initial scraping was carried out using the `Selenium` package for Python (Muthukadan, 2018) in combination with `GNU wget` (Nikšić, 2020). The structure of Folketingstidende.dk is exploited to compile a list of URL links of pdfs-to-be-downloaded. Along with the list of links, metadata for each pdf is collected and stored. We used `wget` along with `GNU Parallel` (Tange, 2018) to download each pdf in the compiled list. Subsequently, the text is extracted from each pdf using the command-line utility `pdftotext` (Cog, 2021), again speeding up the process using `GNU Parallel`. The result is then one text document per pdf. To use this in a meaningful way, the text needs to be segmented into speeches with accompanying speaker names. All transcripts are structured alike, where speeches in the transcripts are preceded by the name of the speaker. The segmentation algorithm uses regular expressions to match speaker names based on a supplied list of members of Folketinget during the period 1990 - 2021. It follows that the quality of this segmentation depends on the provided list of members. As an extension to the original scripts made by Malte Laus, we made sure to include "personskifte", or substitutes, in the list of members of Folketinget. *Substitutes* refer to the replacements of members of Folketinget who for any reason leave Folketinget during their election period. If not provided, the names of the substitutes are left unmatched.

2.2 Preprocessing of the data

(Morten) The result of the segmentation algorithm is a list of delimited text documents that are loaded into `RStudio` (RStudio Team, 2020) for further preprocessing. Each row consists of the matched speaker name, the speech produced by the speaker, the reason for the match, mostly due to the algorithm recognising a speaker name, a unique row identifier and a pdf identifier. The loaded data is joined with the metadata collected in

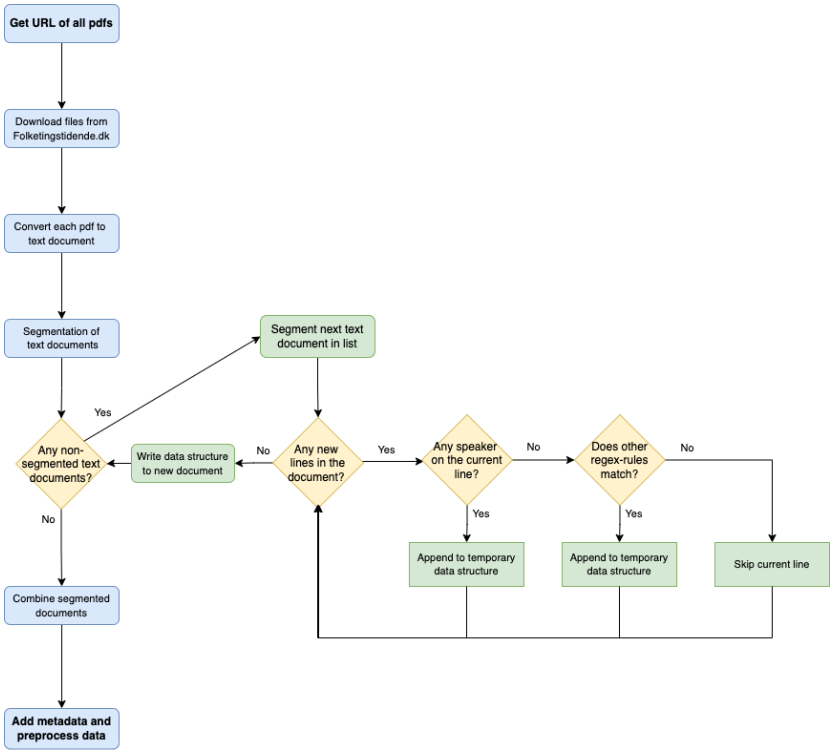

Figure 2: Transformation of pdf transcripts to machine-readable data.

the first step, which contains the date of the parliamentary debate. Before joining the segmented data with metadata about party affiliation, the segmented data is cleaned. This cleaning process involves removing NAs and dropping documents where no speaker names were matched. Data on party affiliation is added based on speaker names and dates. For titles such as "Formand" (chairman), "Udenrigsminister" (Minister of Foreign Affairs) etc., we used a historical list of ministers, their titles and the responding election periods, and subsequently added the name to the data based on title and date. Each row in the resulting data frame is henceforward referred to as a document, and contains speaker name, party affiliation, date, pdf ID, row ID and lastly the words spoken by the speaker.

Following the joining together of data and metadata we made some more influential preprocessing decisions. Firstly, We removed groups with low sample size, i.e. Slesvigsk Parti, Venstresocialisterne, parties from Faroe Islands and Greenland as well as all non-attached members. Secondly, we remove all documents with fewer than 50 words as done in Goet, 2019. This is done to make sure there is sufficient input to measure any meaningful position. Lastly, we remove all pdf IDs from which fewer than 10 instances of speech were detected. This is done to ensure the Wordshoal algorithm has sufficient input to work properly. By inspecting the preprocessed data we found several mistakes regarding the assignment of speech when a chairman was involved. The chairman typically introduces other speakers, and the subsequent speech is incorrectly assigned to the chairman. As a consequence we removed all speeches attributed to the chairmen, both because of this issue, but also because the role of the chairman in Folketinget is to ensure that the legislative process is conducted in accordance with the constitution, and therefore are not a party member giving political speeches in the parliament. The resulting data frame consists of 520.409 rows, each consisting of name, speech as well as time data and party affiliation. Below are some visualisations of the preprocessed data.

As seen in figure 3, number of speeches are obviously correlated with party size. Socialdemokratiet and Venstre, being the two biggest political parties in Folketinget, have produced close to 40% of all speeches in the period 1990-2021. Besides party contribution it might be concerning if speech length per member varied throughout the period. Figure 4 shows a downwards trend, as speech length becomes increasingly smaller. However, the mean speech length is more or less stable, varying between approximately 400 and 250 words. Figure 5 shows an upward trajectory of average number of speeches per person, meaning the number of data points increase as a function of time. Thus, the data indicates that throughout the years the number of speeches has increased while the length of the speeches has decreased.

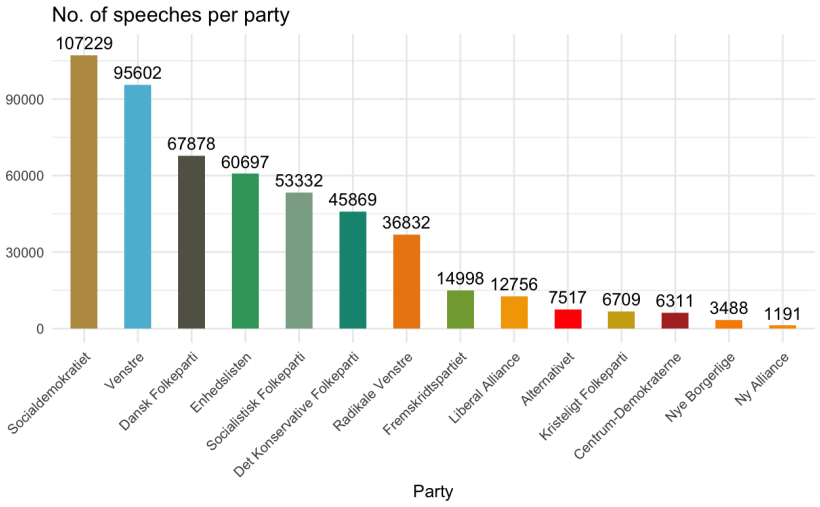

Figure 3: Number of speeches per party (accumulated).

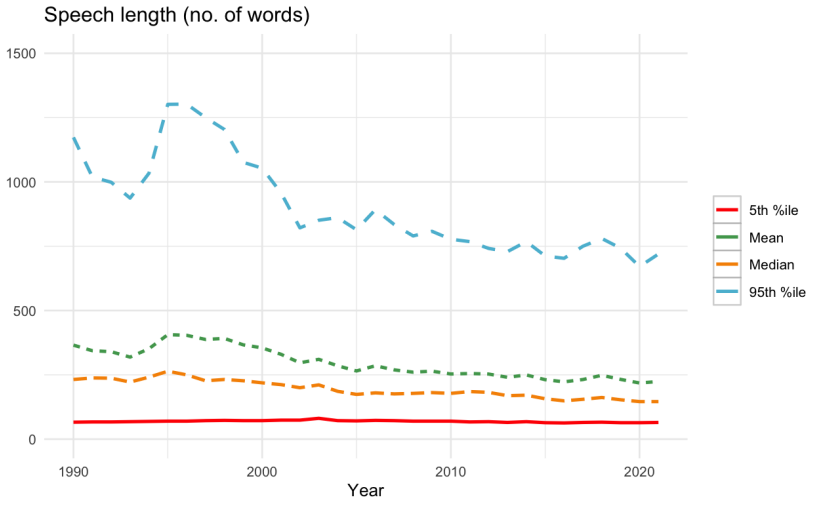

Figure 4: Speech length across years, defined as number of words.

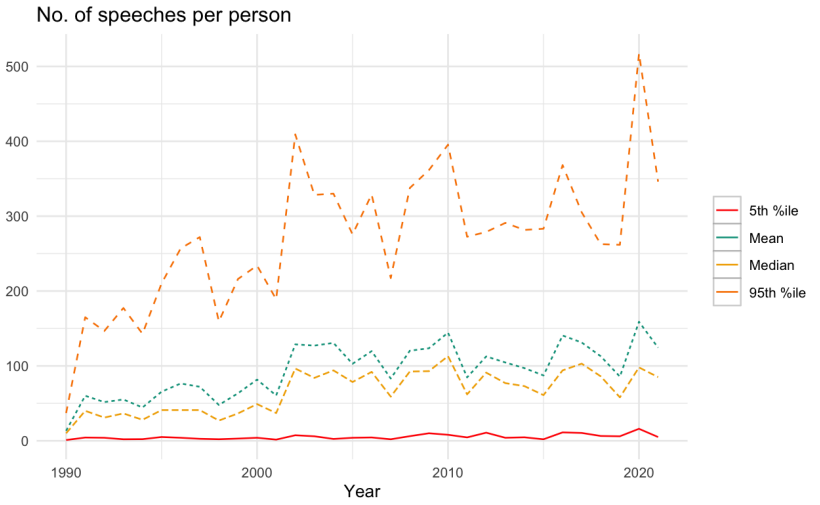

Figure 5: Number of speeches per person across time.

2.3 Validation of the text segmentation

(Gustav) As mentioned some text was assigned incorrectly, and even though we remove some of the more prominent cases, there will obviously still be incorrect assignments. To account for this, we have manually assessed the quality of the data. There are two main sources of error. Either, the segmentation algorithm incorrectly splits the text, or the metadata causes wrong speech assignment. We assess the quality by sampling 200 documents stratified by year and calculating the percentage of correct splits and assignments. The result of the quality assessment shows that approximately 76% of all speeches were assigned and split correctly. The most prominent source of error is incorrect splitting. All the data points in the remaining 24%, were all split incorrectly, with some of them simultaneously being assigned incorrectly. The splitting errors vary in cause, but wrong splitting most frequently happens due to flawed metadata. For example, the title "Fjerde næstformand" (fourth vice chairman) went unmatched on several occasions. Another common error is flawed pdf-to-text transformation, as pdftotext struggles with the formatting of the transcripts.

2.4 Wordshoal

(Morten) The Wordshoal analysis was performed in `R` (R Core Team, 2020). The authors and inventors of the Wordshoal method (Peterson and Spirling, 2018) published along the publication of their article a package that enabled easy replication of their method. Their package `Wordshoal` (B. Lauderdale, 2021) is an extension to the `quanteda.textmodel` framework that is designed for quantitative analysis of textual data (Benoit et al., 2021). The Wordshoal model is implemented for each month in the data set. This means that Wordfish estimates are generated for each meeting in a month and then the Wordfish estimates are scaled across all debates of that month. Note that some days contain two meetings in Folketinget, whereas others only contain one. The model then returns an estimated position for each speaker in that month, as an aggregate of their positions per meeting. By looping through all months in the data set, estimates for all speakers for each month are obtained. Before running the model all punctuations, numbers, symbols and stopwords are removed from the speeches. The list of stopwords were gathered from a list created by the Github user "berteltorp" (Torp, 2021).

2.5 Design of parameters

(Gustav) In order to test H1 and H2 we need estimates of how distinct the parties are from each other and estimates of how assimilated the parties are internally over time. The Wordshoal analysis provides the sufficient estimates to generate both parameters.

The inclusion parameter (i.e. how alike members of the same party are) is interpreted as the standard deviation of the estimated positions for members of a party. Thus a high standard deviation of estimated positions means that a party has low inclusion (i.e. the party members' positions are very spread out), and conversely a low standard deviation of the estimated positions means that the party has high inclusion (i.e. the party members' positions are very similar). In order to put the inclusion measure on a scale from 0 to 1, we divided all inclusion data points with the maximum inclusion value and then subtracted the scores from 1. This made all inclusion values range from a scale from 0 to 1 in which high values represented high levels of inclusion and vice versa.

The distinctiveness parameter (i.e. the positional distance between parties) is obtained through a two-step process. First, we calculate a similarity score. This is defined as the overlap between a specific party's distribution and other parties' distributions. To calculate the overlap score the `overlapping` package for R was used (Pastore, 2018). The advantage of using this package to calculate the overlap metric is that it does not require any distributional assumption to be met (Pastore and Calcagnì, 2019). The overlap metric is calculated from density distributions which is appropriate for the data used in this study as it does not always follow a specific distribution. Estimates are returned for the overlap between all distributions of all parties for a specific month. Then, to gain an estimate of how differentiated the parties are rather than how alike they are, the overlap score is subtracted from one. To give a summarised value for the distinctiveness for each party, the mean of the three nearest parties are calculated and interpreted as the distinctiveness of a specific party. Thus, each party gains an estimate of how differentiated it is from the other parties.

By formalising the parameters as above, we end up with an inclusion score and a distinctiveness score for each party for each month. It is then possible to investigate whether earlier levels of inclusion and distinctiveness predict future levels of inclusion and distinctiveness and thereby test ODT.

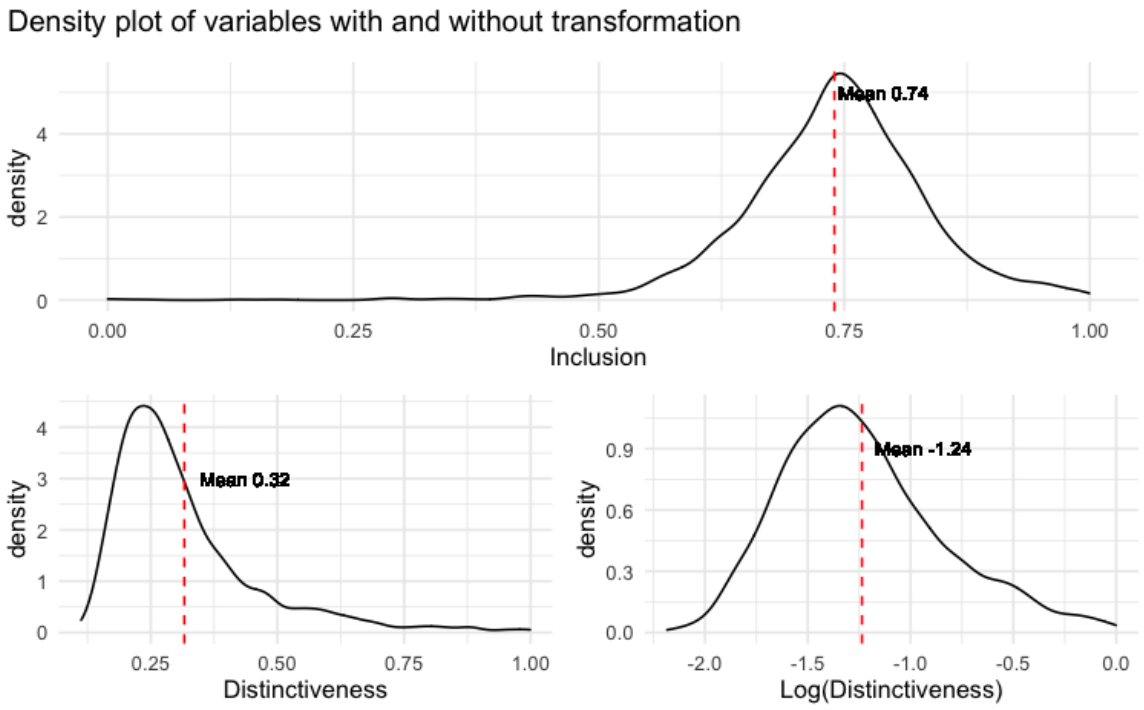

Figure 6: Density plots of outcome variables with and without logarithmic transformation.

2.6 Statistical analysis

(Morten) To test H2 and H3 a vector autoregressive (VAR) model with two variables (k = 2) was fitted within a Bayesian framework. A VAR model is characterised by an outcome variable that is predicted by earlier values of itself and earlier values of other variables (Lütkepohl, 2013). Earlier values are in the literature usually referred to as lagged values, which is also what they will be referred to in the remainder of this paper. A VAR model with k = 2 have the following general structure:

$$Y_t = \beta_{1,0} + \beta_{1,1}Y_{t-1} + \beta_{1,2}Y_{t-2} + \ldots + \beta_{1,p}X_{t-p} + \gamma_{1,1}X_{t-1} + \ldots \gamma_{1,p}X_{t-p} + \epsilon_{1t}$$

$$X_t = \beta_{2,0} + \beta_{2,1}X_{t-1} + \beta_{2,2}X_{t-2} + \ldots + \beta_{2,p}Y_{t-p} + {}_{2,1}\, Y_{t-1} + \ldots \gamma_{2,p}Y_{t-p} + \epsilon_{2t}$$

Where p is the order of the autoregressive model, i.e. the number of lagged values included in the model and X and Y are the two included variables. We build a model with p = 1, as we are interested in causality rather than forecasting. A model with p =1 is a better model for doing causal inferences, as if the value of a variable lagged two time periods influences the present value, this causality will be captured by including the variable lagged one time period (McElreath, 2020).

2.7 Identifying a likelihood function

(Gustav) Before commencing an analysis an appropriate likelihood function for the outcome variable is determined. Through visual inspection of the distributions of inclusion and distinctiveness it was found that distinctiveness was best described with a log-normal distribution and inclusion was best described with a normal distribution. Distinctiveness was described with a log-normal distribution as the logarithm of inclusion approximated normality. The outcome variables are visualised in figure 6 in which distinctiveness before and after transformation is illustrated.

2.8 Model design

(Morten) In order to test H1 and H2 we test the following VAR models with k = 2 and p = 1. The variables included in the model are inclusion and distinctiveness.

$$inclusion_t \sim \beta_{1,0} + \beta_{1,1}inclusion_{t-1} + \beta_{1,2}distinctiveness_{t-1} + \epsilon$$

$$distinctiveness_t \sim \beta_{2,0} + \beta_{2,1} distinctiveness_{t-1} + \beta_{2,2} inclusion_{t-1} + \epsilon$$

The models were run using `brms` (Bürkner, 2017), in which context the two models took the following form:

$$(M1) \;\; inclusion_t \sim 1 + inclusion_{t-1} + distinctiveness_{t-1} + (1|Party)$$
$$(M2) \;\; distinctiveness_t \sim 1 + distinctiveness_{t-1} + inclusion_{t-1} + (1|Party)$$

We build a linear mixed effects model with a random intercept to account for the varying positions of each party. We are interested in the relationship between past and present values of inclusion and distinctiveness, and variance caused by differing default positions of each party is therefore not of interest. The model has n intercepts, where n is the number of parties included in the model.

2.9 Defining priors

(Gustav) The priors for μ and σ are based on summary statistics of the data. NAs and zeroes were omitted from the aggregated dataset. For M2 the distinctiveness variable was log-transformed, whereafter the mean and standard deviation of the outcome variable was calculated and implemented as priors. The same procedure was used for M1 except inclusion was not log-transformed. As for the beta estimates we have implemented a sceptical prior where we expect no effect, with a mean of zero and standard deviation of 0.2. In other words, we expect an effect size within reasonable limits, i.e. 0 +/- 0,4. The prior is identical for both estimates, i.e. $distinctiveness_{t-1}$ and $inclusion_{t-1}$. The prior for the varying intercept was calculated by grouping the data by party, then calculating the mean of inclusion for each party, and lastly calculating the mean and sd of party means which were used as priors. The final priors used are shown next.

$$(1) \;\; inclusion \sim Normal(\mu, \sigma)$$
$$(2) \;\; ln(distinctiveness) \sim Normal(\mu, \sigma)$$

Priors for (1)	Priors for (2)
μ ∼ Normal(0.74, 0.2)	μ ∼ Normal(-1.24, 0.2)
σ ∼ Normal(0.09, 0.2)	σ ∼ Normal(0.39, 0.2)
b ∼ Normal(0, 0.2)	b ∼ Normal(0, 0.2)
sd ∼ Normal(0.74, 0.11)	sd ∼ Normal(0.57, 0.05)

2.10 Model fitting and quality checks

(Morten) Fitting of the model was carried out with Hamiltonian Monte Carlo samplers, employing four parallel chains, each with 2000 iterations. An adept delta of 0.95 was applied to assist the model with divergent transitions, and a max tree depth of 20 was implemented. We found no divergences and all Rhat values equalled 1, indicating the chains have explored the parameter space effectively. Through visual inspection of the Markov chains we found no obvious issues regarding the sampling process. Prior predictive checks were conducted to check predictions made from the priors. The sampled values for the priors were deemed to resemble the actual values of inclusion and distinctiveness in a satisfactory way. Comparison of the prior and posterior showed that the posterior had updated its prior beliefs from the data. Plots of the diagnostics of the sampling process can be seen in Appendix A for M1 and Appendix B for M2.

3 Results

(Gustav) The first model with inclusion as outcome variable yielded the following results. For lagged values of inclusion a credible negative effect was found ($\beta_{1,1}$= -.02, SE = 0, 95% CI [-.04, -.00], ER = 46.06). A credible effect was also found for lagged values of distinctiveness ($\beta_{1,2}$= .03, SE = 0, 95% CI [.01, -.06], ER = 44.98). The positive beta estimate for lagged values of distinctiveness, indicates a positive association between the outcome variable inclusion and lagged distinctiveness. Both the effects have evidence ratios around 45, which indicates that the model found it 45 times more plausible that the effect is different from null than that the effect does not exist. The effect of lagged distinctiveness is visualised in figure 7.

The second model with distinctiveness as outcome variable yielded the following results. Lagged values of distinctiveness was found to have a credible effect on distinctiveness ($\beta_{2,1}$= .4, SE = 0, 95% CI [.31, .50], ER = Inf). However, no credible effect was found for lagged values of inclusion ($\beta_{2,2}$= .00, SE = 0, 95% CI [-.06, .07],

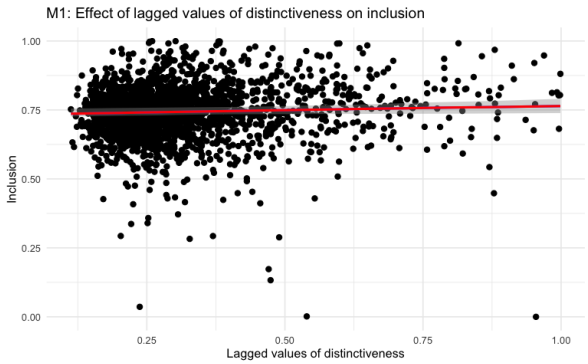

Figure 7: Effect of lagged differentiation on inclusion.

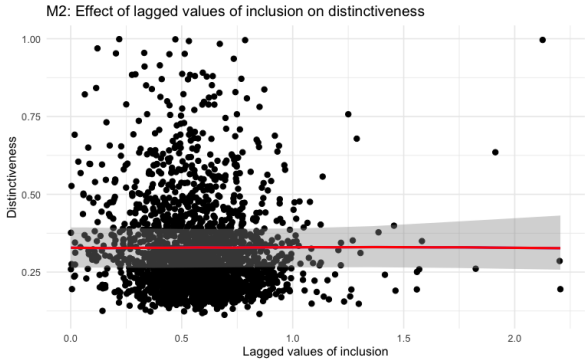

Figure 8: Effect of lagged inclusion on distinctiveness.

ER = 0.89). That no effect was found for lagged values of inclusion means that the model did not find any evidence that the variable provided new information on inclusion when lagged values of inclusion were known. The effect of lagged inclusion is visualised in figure 8.

4 Discussion

(Morten) Model 1 (M1) was designed and conducted to test hypotheses 1 (H1). To summarise, the aim of H1 was to investigate whether inclusion was positively associated with lagged values of distinctiveness. The outcome variable of M1 was inclusion and the variables included in the formula were lagged values of inclusion and lagged values of distinctiveness. The results of our first model showed lagged values of both inclusion and distinctiveness had a credible effect on current levels of inclusion. Furthermore, the estimate for lagged values of distinctiveness was positive. This means that the results of M1 does not provide any evidence that suggests H1 should be discarded. Conversely, the results correspond to the predictions made from H1, that there is a positive association between inclusion and lagged values of distinctiveness.

(Gustav) Model 2 (M2) was designed and conducted to test hypothesis 2 (H2). The aim of H2 was to investigate whether the relationship between inclusion and distinctiveness was interdependent. In other words, whether they were both drivers for each other. Thus, H2 was a reversed version of H1. More specifically, H2 tested whether distinctiveness was positively associated with lagged values of inclusion. Therefore, the outcome variable of M2 was specified as distinctiveness and the variables included as predictors were lagged values of distinctiveness and lagged values of inclusion. The results of M2 yielded a credible positive effect for lagged values of distinctiveness, but no credible effect for lagged values of inclusion. Thus, the model did not find that including lagged values of inclusion provided new information that helped explain the outcome variable distinctiveness which was not already explained by lagged values of distinctiveness. The results of M2 does thereby not provide evidence that corresponds with H2, hence no positive association was found between distinctiveness and previous levels of inclusion.

(Morten) M1 yielded that inclusion was explained by lagged values of both inclusion and distinctiveness while M2 yielded that distinctiveness was only explained by lagged values of distinctiveness and not lagged values

of inclusion. That the effect of the predictor of interest is only present in one of the models suggests that there might be a hierarchical ranking of the two predictors in which one of the predictors drives the other. That the analysis was of time series data, further supports this as it implies the predictor of interest is bound to the outcome variable by time. This means that reversing the relationship between the outcome variables and predictors does not necessarily yield the same results as if it is modelled independent of time by just predicting inclusion from distinctiveness. As distinctiveness explained inclusion and inclusion did not explain distinctiveness, the results of the models suggest that distinctiveness drives inclusion and not the opposite.

(Gustav) Another function of including lagged values of the outcome variable as a predictor in the model, was to ensure that the model accounted for natural oscillations in inclusion or distinctiveness and did not attribute these oscillations as an effect of the predictor of interest. Furthermore by modelling the lagged values of the outcome variables, we could account for a confounder introduced in the design of our parameters. Inclusion was formalised as the standard deviation of Wordshoal estimates for a party, while distinctiveness was formalised from scores of overlap. Due to the parameter design, we would expect a positive association between the two variables, since it is to be expected that as the standard deviation (inclusion) of a party increases it will fill up more space in the parameter space. By filling up more space in the parameter space it will simultaneously have bigger overlaps with other parties' distributions (distinctiveness). By modelling the variables in a vector autoregressive framework, we ensure that the lagged values of the outcome variable accounts for variance due to natural oscillations and the confounder of the parameter design. Meanwhile, the estimates of the predictor of interest only tells whether the predictor adds any new information to the model which is not already explained in the lagged values of the outcome variable. Thereby, the estimates of the predictor of interest should provide information on the effect of interest rather than other confounders.

In an exploratory analysis, an interaction effect was added to both models. The interaction yielded credible effects in both models suggesting that the relationship between distinctiveness and inclusion might not be linear. Since modelling this was not within the scope of the study, the effect will not be reported and interpreted on, but simply be mentioned to encourage future research into unravelling the implications of a possible interaction effect.

4.1 Corroboration with Optimal Distinctiveness Theory

(Morten) In the following paragraphs the results of the models will be put into a theoretical context in which correspondences and discrepancies with the existing literature will be highlighted. As mentioned, results pertaining to H1 seemed to support the notion of positive association between previous level of distinctiveness and current level of inclusion. This corroborates the ODT, as high values of distinctiveness will be followed by high values of inclusion. In other words, this indicates that actors are adjusting their level of inclusion within their group to compensate for the level of distinctiveness between groups, to maintain an equilibrium between the two opposing needs. The results pertaining to H2 tells a slightly different story. When accounting for the autocorrelation of distinctiveness, previous levels of inclusion do not predict current levels of distinctiveness. According to ODT, we would expect a positive effect of the previous levels of inclusion on current levels of distinctiveness. The missing effect indicates that the groups in Folketinget do not adjust their distinctiveness between groups according to previous levels of inclusion within their groups. Thus, adjustment of the level of distinctiveness is not used as a strategy of the groups in Folketinget to maintain equilibrium. Accordingly to ODT, groups can either adjust their level of inclusion or their level of distinctiveness to maintain equilibrium between the two needs. The results of H1 and H2 suggest that groups in Folketinget only adjust their level of inclusion within the group to maintain equilibrium between the two needs.

(Gustav) There might be an interesting inference to be made when considering the results of the two models together. The results seem to suggest that distinctiveness is a driver of behavioural change in Folketinget, in contrast to inclusion which does not seem to drive any behavioural change. When Brewer, 1991, and later Leonardelli et al., 2010, published their accounts of ODT, no hierarchy of needs was established. For this reason, it is reasonable to expect inclusion and distinctiveness to have an equal influence on behaviour. This expectation is seen in our hypotheses. However, both authors emphasise the strength of each need depends on cultural norms and individual mentality. In other words, perceived distinctiveness is directly related to the social arena of the group, i.e highly context dependent. Our results seem to indicate that positional differences between parties are not influenced by previous levels of inclusion. The lack of effect of inclusion on distinctiveness might be explained by the specific arena of choice. Folketinget was chosen as the arena of study and as Folketinget is a political scene, it might be expected that there are other influential variables on the position of the parties than the level of inclusion within the party. In a political scene as Folketinget, the party profile is important to maintain. For example, when Nye Borgerlige entered Folketinget, voters quickly realised their immigration politics were similar to that of Dansk Folkeparti. Resultantly, voters focused on immigraton politics voted for Nye Borgerlige, as the voter can be assured the new party will fight for their core area. This

in turn prompts Dansk Folkeparti to demarcate themselves from Nye Borgerlige. This example illustrates that political interests might have an influence on the position of the party. As inclusion within parties was not found to have any effect on distinctiveness between the parties it might be because there are political interests that are more influential on the position of the party. However, this does not entail that the parties do not behave to maintain equilibrium between the two needs. It entails that the groups within Folketinget adjust their level of inclusion rather than their level of distinctiveness to maintain the equilibrium, as there might be other matters of interest of the group like political interests that need to be maintained through the position of the group.

4.2 Methodological considerations

4.2.1 Preprocessing and its implications

(Morten) As displayed in figure 1, the process of transforming pdf transcripts to machine-readable data contains multiple steps. Each of these steps might distort the input to the model, in which case any inferences drawn should be regarded with some suspicion. The early processing of the input can be conceptualised as an extension of the statistical models, due to its impact on the estimates and inferences drawn. Consequently, it seems appropriate to consider the transformation steps taken, and their potential impact. In this analysis there are two important transformations that might lead to flawed input; transformation of pfds to text documents and segmentation of the text documents. As mentioned, the command-line utility `pdftotext` (Cog, 2021) was used for the transformation from pdfs to text documents. The transformation caused several issues revealed in the manual validation of the data. In some cases, the transformer did not correctly interpret the two-column format of the transcripts. The issue was particularly associated with old pdfs, and was part of the reason for discarding transcripts from the period 1953-1989 as the error rate increased for old transcripts. An idea for improval is to implement other algorithms for pdf-to-text transformations and hopefully enable the researchers to extend the period of analysis from 1990-2021 to 1953-2021

(Gustav) Another source of error are the headers and footers of the pdfs, as the text contained within them is frequently interpreted as speech. This was partially solved through preprocessing based on regular expressions, however the issue is still present. As for the segmentation of each text document, the quality of the process is highly dependent on solid metadata.. The transcriptions are formatted with a preliminary note containing the member's name and title followed by the speech of the member. This note occasionally contains the full name of the speaker, but oftentimes it is some combination of title, first name and/or last name(s). The segmentation algorithm used in this analysis was revealed to miss some combinations of these which led to cases of incorrect segmentation. The combinations that were missed were often caused by incomplete metadata. It was discovered some members of Folketinget were not matched, as they were substitutes for an elected parliament member and therefore not included in the list of members in Folketinget. Furthermore, the list of official titles within Folketinget proved imperfect, especially for titles pertaining to the early days of transcription. A considerate improval of the segmentation of the text documents could be achieved by a thorough revision of the metadata.

4.2.2 Implications of using behavioural measures

(Morten) In this study we investigate group structures through intergroup behaviour with behavioural measures. This comes with some assumptions. Firstly, we implicitly equal perceived distinctiveness vs. actual distinctiveness. In other words, there is no phenomenological rapport given, so it is assumed the members of Folketinget perceive the same group structure as the one we measure. For example, we expect that a member of Venstre will perceive the same positional similarity of intragroup members as measured. Secondly, the needs are being operationalized in a certain way. In this analysis, inclusion is seen as deindividuation, and it is formalised as a party's positional similarity. Distinctiveness is operationalized as the inverse of the overlap between the party's and the positions of the three closest parties. The choice of only including the three closest parties is arbitrary and can be changed, but besides this choice the operationalization of distinctiveness seems fairly intuitive. Finally, we take for granted human beings assume their role as party members when they act in Folketinget. This assumption may hold true in general, with some inconsistencies arising when the human beings behind the party members do not immerse themselves fully. The point is the predictions being made are sensible only insofar as the actors involved behave as members of the political groups.

4.2.3 Natural language as a measure of intergroup behaviour

(Gustav) The entire study at hand rests on the assumption that it is possible to make inferences on intergroup behaviour from word usage. All the variables used in the analysis are obtained through transformations of natural text. However, that word use is a representative measurement for intergroup behaviour is not a matter of course. It has long been theorised that language influences cognition and vice versa. One of the first theories

on this was the Sapir-Whorf hypothesis, which stated that language shapes cognition (Boroditsky, 2011; Kay and Kempton, 1984). Past decades, numerous studies have demonstrated how language shapes the way we think. Studies have shown that changing how people speak changes how they think (Kroll and Bialystok, 2013), or that teaching people new words for colours enhances their ability to discriminate between colours (Thierry et al., 2009). But, does cognition also alter the way we speak and is the way we speak a representative measure for how we think? Within the field of cognitive linguistics it is generally assumed that language should reflect what is known about human cognition from other cognitive and brain sciences (Evans, 2012). Thus, it should be safe to assume that the cognitive process of social identity investigated in this paper should be reflected in language. However, if language can reflect cognitive processes, it is most likely not consistently representing a specific type of cognitive process in all language use. The specific cognitive process that language reflects in language is most likely highly dependent on the context and motivation of the speaker. Nonetheless, all language used in this study is gathered from the same context, Folketinget, in which the common task of processing bills are consistent throughout sessions (cf. paragraph from introduction). Although the task and context could be further divided into more specific subcontexts and subtasks, we argue that the context and task of the language used in this study are somewhat consistent, why the behaviour represented in the language should also be somewhat consistent.

4.2.4 External sources of variation

(Morten) Whether the behaviour reflected in the language in Folketinget is motivated by ODT is difficult to determine. In this study we have assumed that the way the parties are structured internally and relatively are solely motivated by the excited needs for either inclusion or distinctiveness. As Folketinget is not a closed system, external factors will inevitably influence the group structure, thereby not exclusively driven by the need for inclusion and distinctiveness. There might be an endless number of such external variables that could have influence on how the members of Folketinget behave. One influential variable could be that members of Folketinget are representatives of the ones who voted for them. It would therefore be expected that members tend to behave in correspondence with the voters' expectations. It should therefore be recognized that there are other drivers of the behaviour measured in the members of Folketinget than the needs for inclusion or distinctiveness.

4.3 Validity of the Wordshoal estimates

(Gustav) As mentioned in the introduction, Wordshoal arose as an attempt to solve the problem of how to account for topic variation in Wordfish analyses. To summarise, it is generally theorised that a hierarchy exists in language variation and if belief variation is to be measured the exceeding levels has to be accounted for. Otherwise the estimates will reflect variation caused by one of the higher levels in the hierarchy. Wordshoal assumes that topic variation can be accounted for by generating Wordfish estimates per session instead of conducting one Wordfish analysis of the whole corpus. B. E. Lauderdale and Herzog, 2016, argued that it could be assumed the topic did not vary within each session and thereby account for topic variation by conducting the analysis per session. This assumption is however very sensitive to the parliamentary system on which the analysis is conducted. The method was developed in the Irish Dail and British Parliament in which a session might be more restricted to a single topic. However, this does not generalise across parliaments. A typical session in Folketinget will typically process more than one bill, that may vary in topic. Therefore, the assumption that topic variation is accounted for when conducting Wordshoal on Folketinget might be violated. A violation of this assumption would imply that the estimates generated in this study would represent topic variation rather than variation caused by different beliefs. As we in this study investigate political parties which are grouped by political beliefs rather than topics, the estimates would not position the parties correctly according to political positions. A violation of the assumption would thereby invalidate the analysis of this study. Future research might attempt to account for topic variation differently. For instance, a topic model like seeded LDA could be conducted on the speeches preceding the analysis. A seeded LDA analysis estimates constituent topics of a speech from prior defined topics with associated words (Jagarlamudi et al., 2012). Thus, seeded LDA could ensure that the topics the speeches are grouped by are meaningful. The LDA scores could then be used to group the speeches by topic and ensure that speeches of different topics were not compared. A Wordshoal analysis could then be conducted on the different topics rather than on different sessions.

(Morten) However, it is difficult to validate the Wordshoal method and the estimates it generates, as there are no objective measures against which the estimates and results can be compared. As an attempt to validate the Wordshoal method, we implemented another previously validated method used to measure political polarisation on the data. We employed the machine classifier mentioned in the introduction section, originally designed by Peterson and Spirling, 2018, and validated by Goet, 2019. Even though our analysis is not concerned with levels of polarisation, Wordshoal estimates have previously been used as a measure of polarisation. If the estimates of the two methods are correlated, Wordshoal generates similar estimates as a validated method,

pointing towards validity of Wordshoal. We implemented the multivariate classifier within the framework of the Python library `scikit-learn` (Pedregosa et al., 2011). The classifier was trained in assigning party labels to speeches. The accuracy of the model is used as a measure of political polarisation, where high accuracy represents a period of high polarisation and low accuracy represents a period of low polarisation. In order to fit the classifier to the data the documents are vectorized using TF-IDF vectorization. We use a logarithmic loss function and balanced class weights to account for the large variance in group size. The classifier was fitted on a yearly basis. As the selected time period consists of 31 years, the output of the fitting procedure is 31 metric values. Cohen's Kappa is a performance metric similar to simple accuracy measures. In Folketinget, the bigger political parties have produced most of the documents, resulting in an increased probability of producing a correct prediction for these parties by chance. This is penalised by Cohen's Kappa, which takes into account the probability of predicting accurately by chance. We compared the 31 Kappa scores with the difference scores obtained through Wordshoal and the overlapper-function, summarised by year. If we assume the classifier represents the real world, and the Wordshoal measures are valid, we expect a positive correlation between the two. A correlation test showed no significant correlation between the estimates of the two methods, r(30) = -.20, p = 0.28. This suggests the Wordshoal estimates lack ecological validity. The estimates are plotted in 9.

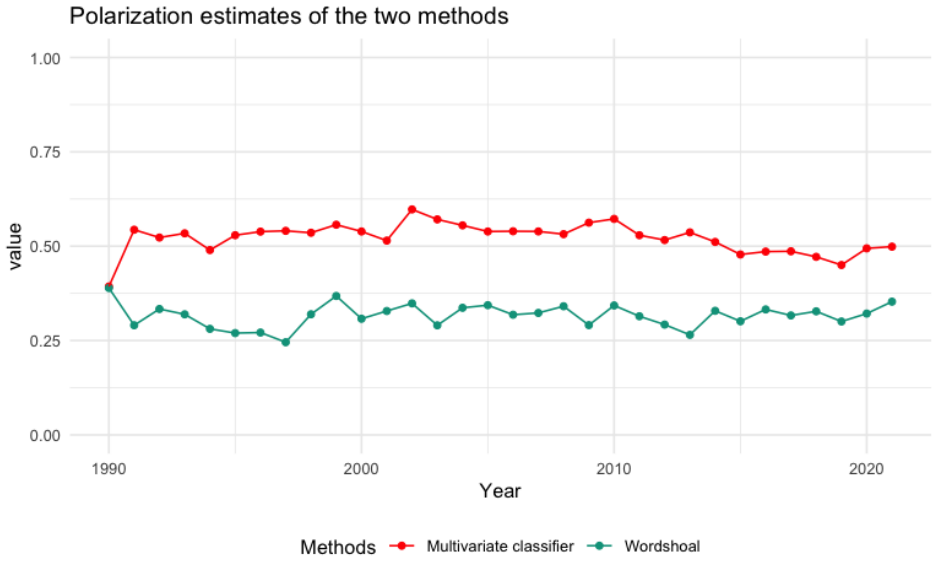

Figure 9: Polarisation estimates for the two different methods.

5 Conclusion

(Gustav) We hypothesised that parties in Folketinget would behave according to the Optimal Distinctiveness Theory (Brewer, 1991; Leonardelli et al., 2010). The results showed a credible effect of lagged values of distinctiveness on current values of inclusion. The opposite was not the case, as lagged inclusion did not predict distinctiveness. The results suggest that groups in Folketinget adjust levels of inclusion according to levels of distinctiveness and not the opposite to maintain equilibrium. The preprocessing of the data is encouraged to be improved to increase validity of the inferences drawn from the data, but is deemed sufficiently valid to suggest the effects of the observed intergroup behaviour. The estimates generated from Wordshoal were validated against estimates obtained from a multivariate classifier. No significant correlation was found between the estimates of the two methods, suggesting a possible lack of validity of the Wordshoal estimates. Using Wordshoal as a method to obtain estimates and use these to measure group behaviour is rather novel. Even though it has been encouraged in previous research (B. E. Lauderdale and Herzog, 2016), it has not been performed yet. We encourage replication studies, either by replicating the analysis of intergroup behaviour according to ODT in Folketinget using other methods or by using the same method in different social arenas.

References

Araujo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on [Publisher: Taylor & Francis]. *Ecological Psychology*, *19*(1), 69–78.

Bem, D. J., & Lord, C. G. (1979). Template matching: A proposal for probing the ecological validity of experimental settings in social psychology. [Publisher: American Psychological Association]. *Journal of Personality and Social Psychology*, *37*(6), 833.

Benoit, K., Watanabe, K., Wang, H., Perry, P. O., Lauderdale, B., Gruber, J., & Lowe, W. (2021). *Quanteda.textmodels: Scaling Models and Classifiers for Textual Data*. https://CRAN.R-project.org/package=quanteda.textmodels
R package version 0.9.4

Böhm, R., Rusch, H., & Baron, J. (2020). The psychology of intergroup conflict: A review of theories and measures [Publisher: Elsevier]. *Journal of Economic Behavior & Organization*, *178*, 947–962.

Boroditsky, L. (2011). How language shapes thought [Publisher: JSTOR]. *Scientific American*, *304*(2), 62–65.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. [Publisher: American Psychological Association]. *Psychological bulletin*, *86*(2), 307.

Brewer, M. B. (1991). The social self: On being the same and different at the same time [Publisher: Sage Publications Sage CA: Thousand Oaks, CA]. *Personality and social psychology bulletin*, *17*(5), 475–482.

Brewer, M. B. (1997). On the social origins of human nature. [Publisher: Blackwell Publishing].

Cog, G.
bibinitperiod. (2021). Pdftotext.

Dubé, L., & Guimond, S. (1986). Relative deprivation and social protest: The personal-group issue. *Relative deprivation and social comparison: The Ontario symposium*, *4*, 201–216.

Dubois, N., & Beauvois, J.-L. (2005). Normativeness and individualism [Publisher: Wiley Online Library]. *European Journal of Social Psychology*, *35*(1), 123–146.

Evans, V. (2012). Cognitive linguistics [Publisher: Wiley Online Library]. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(2), 129–141.

Goet, N. D. (2019). Measuring Polarization with Text Analysis: Evidence from the UK House of Commons, 1811–2015 [Publisher: Cambridge University Press]. *Political Analysis*, *27*(4), 518–539. https://doi.org/10.1017/pan.2019.2

Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: Stereotype threat and distance in interracial contexts. [Publisher: American Psychological Association]. *Journal of personality and social psychology*, *94*(1), 91.

Jagarlamudi, J., Daumé III, H., & Udupa, R. (2012). Incorporating lexical priors into topic models. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 204–213.

Jetten, J., Spears, R., & Postmes, T. (2004). Intergroup distinctiveness and differentiation: A meta-analytic integration. [Publisher: American Psychological Association]. *Journal of personality and social psychology*, *86*(6), 862.

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? [Publisher: Wiley Online Library]. *American anthropologist*, *86*(1), 65–79.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition [Publisher: Taylor & Francis]. *Journal of cognitive psychology*, *25*(5), 497–514.

Lauderdale, B. (2021). *Wordshoal: Quanteda textmodel extensions for the wordshoal textmodel*. http://github.com/kbenoit/wordshoal
R package version 0.4

Lauderdale, B. E., & Herzog, A. (2016). Measuring Political Positions from Legislative Speech [Publisher: Cambridge University Press]. *Political Analysis*, *24*(3), 374–394. https://doi.org/10.1093/pan/mpw017

Leonardelli, G. J., Pickett, C. L., & Brewer, M. B. (2010). Optimal distinctiveness theory: A framework for social identity, social cognition, and intergroup relations. *Advances in experimental social psychology* (pp. 63–113). Elsevier.

Lütkepohl, H. (2013). Vector autoregressive models. *Handbook of research methods and applications in empirical macroeconomics*. Edward Elgar Publishing.

Lutz, A., & Thompson, E. (2003). Neurophenomenology Integrating Subjective Experience and Brain Dynamics in the Neuroscience of Consciousness. *Journal of Consciousness Studies*, *10*(9-10), 31–52. https://www.ingentaconnect.com/content/imp/jcs/2003/00000010/f0020009/art00004

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Vol. 2nd ed). Chapman; Hall/CRC. https://doi.org/10.1201/9780429029608

McGuire, W. J., & McGuire, C. V. (1988). Content and process in the experience of self. *Advances in experimental social psychology* (pp. 97–144). Elsevier.

Muthukadan, B. (2018). Selenium with Python.

Nikšić, H. (2020). GNU wget.

Olivares, F. A., Vargas, E., Fuentes, C., Martínez-Pernía, D., & Canales-Johnson, A. (2015). Neurophenomenology revisited: Second-person methods for the study of human consciousness. *Frontiers in Psychology*, *6*, 673. https://doi.org/10.3389/fpsyg.2015.00673

Pastore, M. (2018). Overlapping: A R package for Estimating Overlapping in Empirical Distributions. *The Journal of Open Source Software*, *3*(32), 1023. https://doi.org/10.21105/joss.01023

Pastore, M., & Calcagnì, A. (2019). Measuring Distribution Similarities Between Samples: A Distribution-Free Overlapping Index. *Frontiers in Psychology*, *10*, 1089. https://doi.org/10.3389/fpsyg.2019.01089

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peterson, A., & Spirling, A. (2018). Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems [Publisher: Cambridge University Press]. *Political Analysis*, *26*(1), 120–128. https://doi.org/10.1017/pan.2017.39

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. http://www.rstudio.com/

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis [Publisher: Wiley Online Library]. *Infancy*, *2*(4), 419–436.

Slapin, J. B., & Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts [Publisher: Wiley Online Library]. *American Journal of Political Science*, *52*(3), 705–722.

Smith, D. W. (2018). Phenomenology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2018/entries/phenomenology/

Snyder, C. R., & Fromkin, H. L. (1977). Abnormality as a positive characteristic: The development and validation of a scale measuring need for uniqueness. [Publisher: American Psychological Association]. *Journal of Abnormal Psychology*, *86*(5), 518.

Snyder, C. R., & Fromkin, H. L. (2012). *Uniqueness: The human pursuit of difference*. Springer Science & Business Media.

Strømberg-Derczynski, L., Ciosici, M. R., Baglini, R., Christiansen, M. H., Dalsgaard, J. A., Fusaroli, R., Henrichsen, P. J., Hvingelby, R., Kirkedal, A., Kjeldsen, A. S., et al. (2020). The Danish gigaword project. *arXiv preprint arXiv:2005.03521*.

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, *56*(65), 9780203505984–16.

Tange, O. (2018). GNU Parallel. https://doi.org/10.5281/zenodo.1146014.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception [Publisher: National Acad Sciences]. *Proceedings of the National Academy of Sciences*, *106*(11), 4567–4570.

Torp, B. (2021). Stopord. https://gist.github.com/berteltorp/0cf8a0c7afea7f25ed754f24cfc2467b

Appendices

A Model 1 quality checks

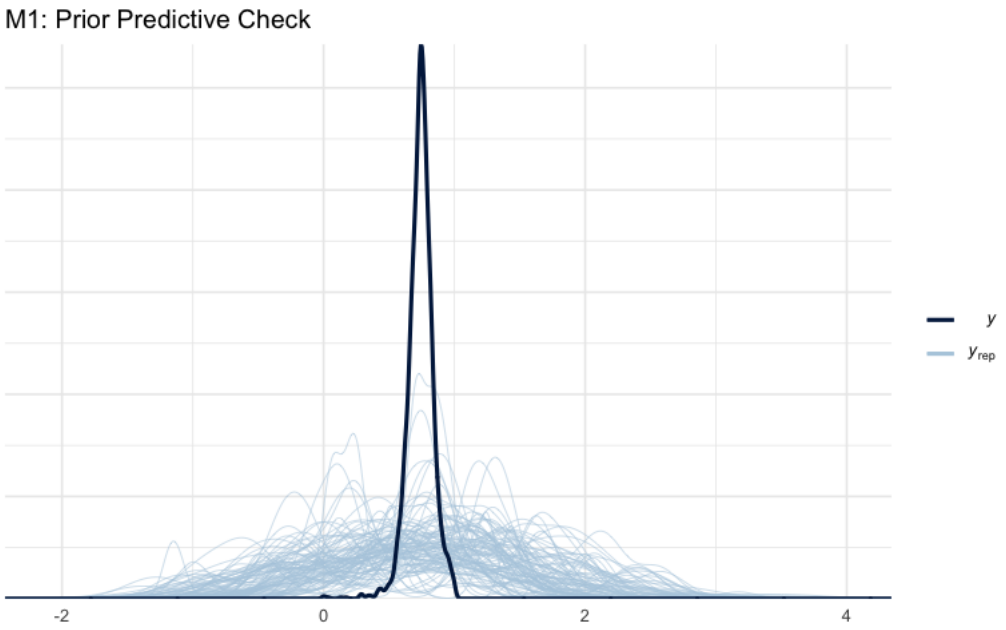

Figure 10: Prior predictive checks for model M1.

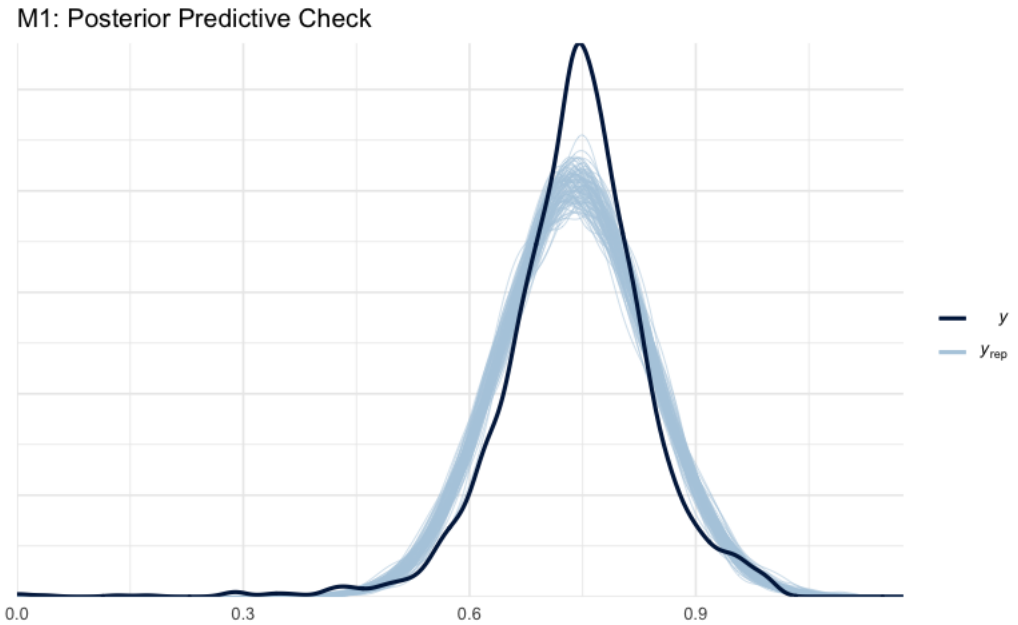

Figure 11: Posterior predictive checks for model M1.

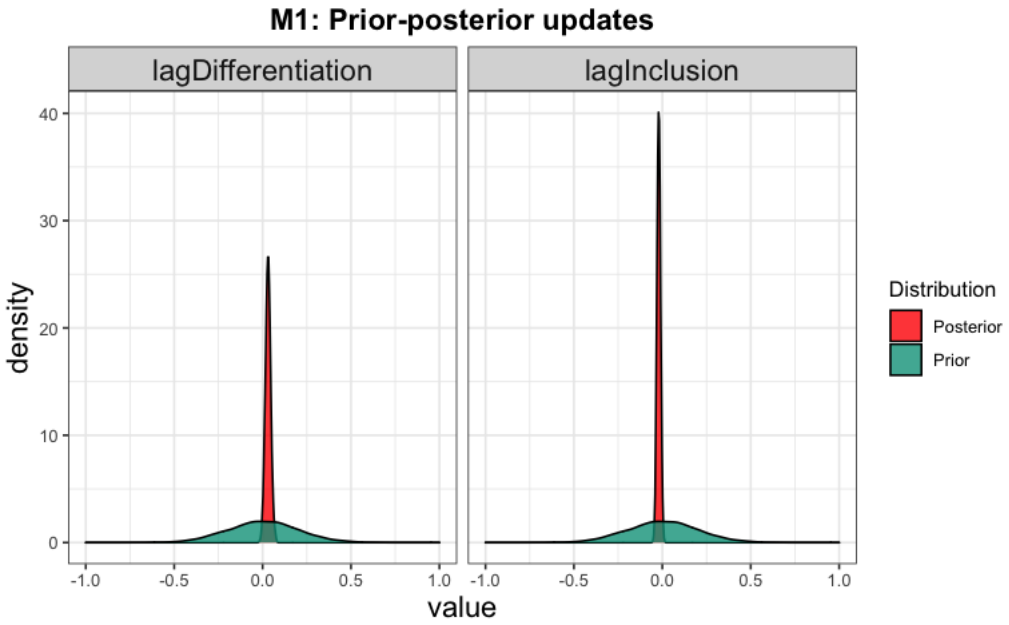

Figure 12: Prior-posterior updates for model M1.

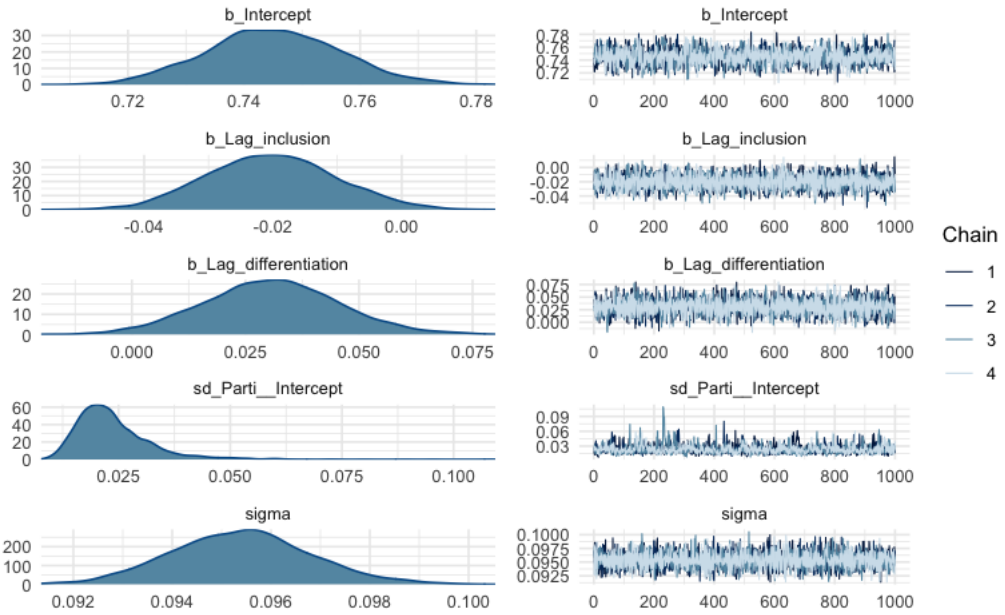

Figure 13: Trace plots for model M1.

B Model 2 quality checks

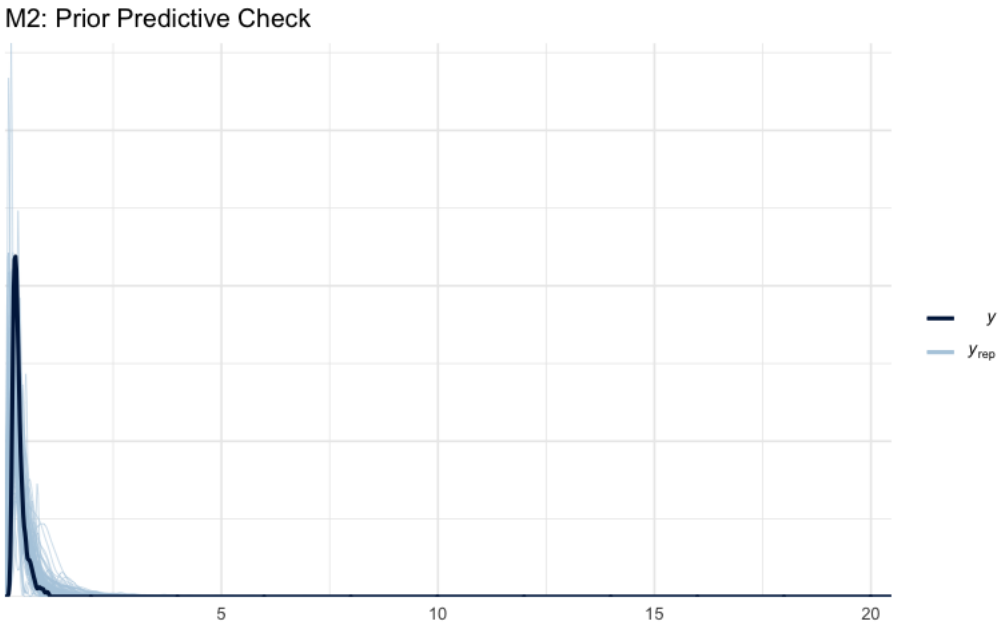

Figure 14: Prior predictive checks for model M2.

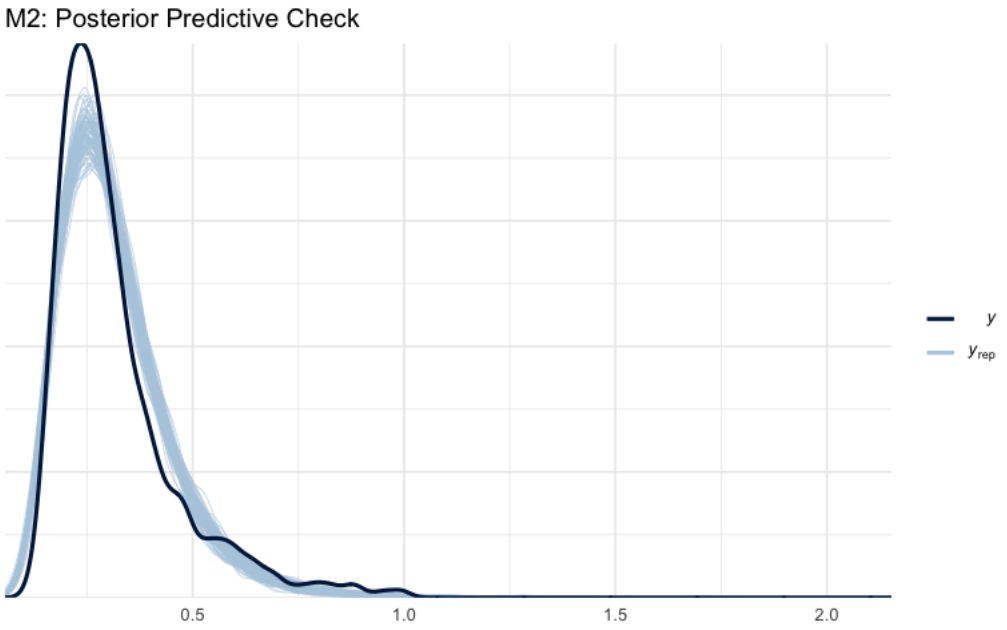

Figure 15: Posterior predictive checks for model M2.

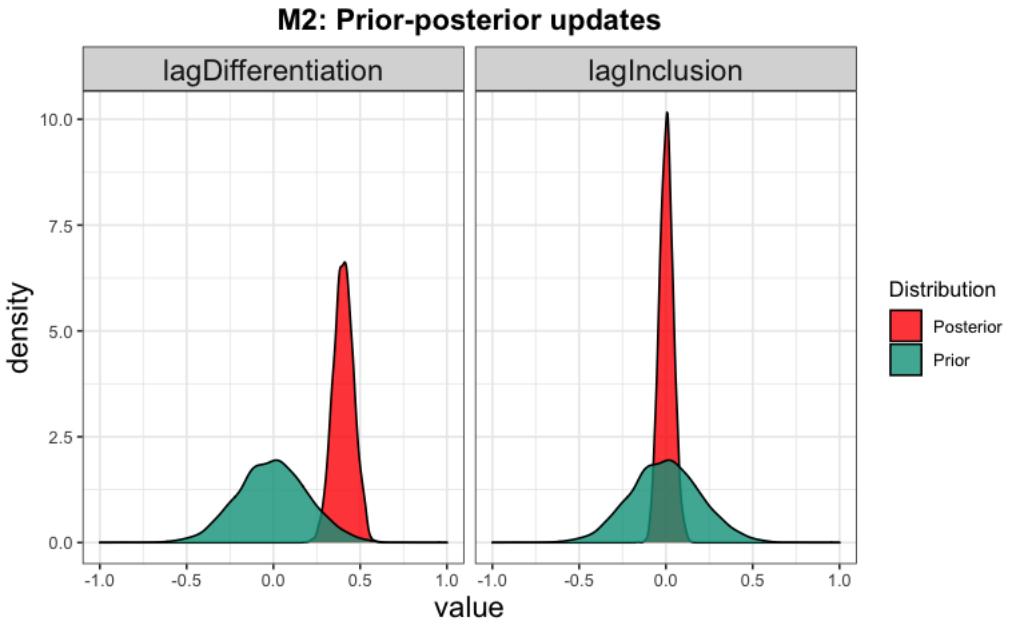

Figure 16: Prior-posterior updates for model M2.

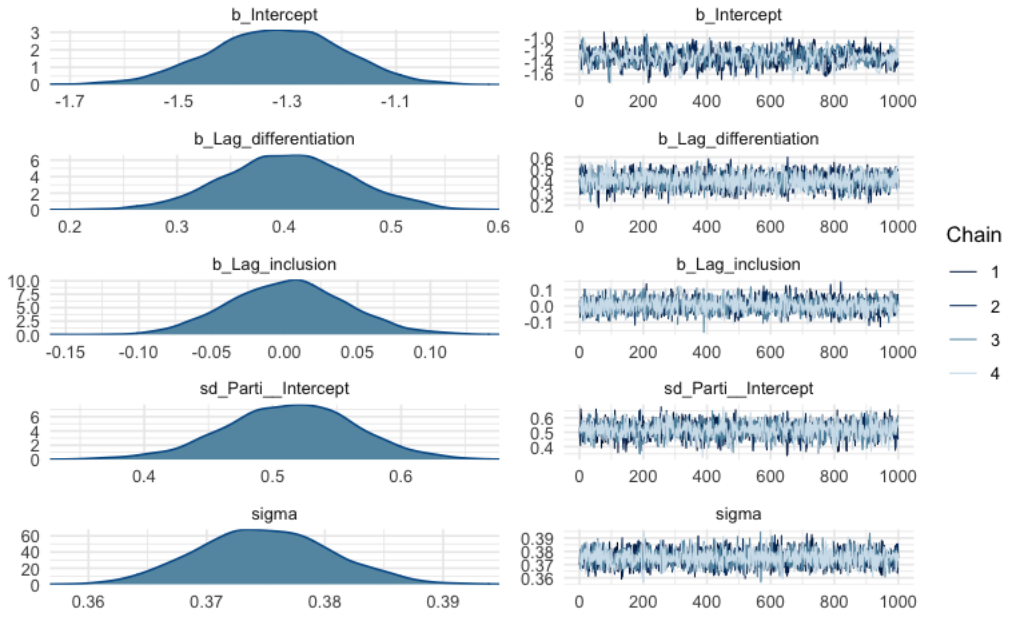

Figure 17: Trace plots for model M2.

C PDF example

4 Tirsdag den 20. november 2018 (L 24)

Det er vedtaget.

Det næste punkt på dagsordenen er:
6) 2. behandling af lovforslag nr. L 24:
Forslag til lov om supplerende bestemmelser til forordning om opstilling af rammer for energimærkning m.v. (produktenergimærkningsloven).
Af energi-, forsynings- og klimaministeren (Lars Christian Lilleholt).
(Fremsættelse 03.10.2018. 1. behandling 10.10.2018. Betænkning 15.11.2018).

Kl. 13:05

Forhandling

Formanden (Pia Kjærsgaard):
Er der nogen, der ønsker at udtale sig?
Da det ikke er tilfældet, er forhandlingen sluttet, og vi går til afstemning.

Kl. 13:05

Afstemning

Formanden (Pia Kjærsgaard):
Ønskes afstemning om ændringsforslag nr. 1-5, tiltrådt af udvalget?
De er vedtaget.

Jeg foreslår, at lovforslaget går direkte til tredje behandling uden fornyet udvalgsbehandling. Hvis ingen gør indsigelse, betragter jeg det som vedtaget.
Det er vedtaget.

Det næste punkt på dagsordenen er:
7) 2. behandling af lovforslag nr. L 38:
Forslag til lov om ændring af lov om almene boliger m.v., lov om leje af almene boliger og lov om leje. (Nye kriterier for udsatte boligområder og ghettoområder, initiativer til udvikling eller afvikling af ghettoområder, skærpelse af anvisnings- og udlejningsregler, ophævelse af lejekontrakt på grund af kriminalitet m.v.).
Af transport-, bygnings- og boligministeren (Ole Birk Olesen).
(Fremsættelse 03.10.2018. 1. behandling 11.10.2018. Betænkning 15.11.2018).

Kl. 13:06

Forhandling

Formanden (Pia Kjærsgaard):
Transport-, bygnings- og boligministeren har meddelt mig, at han ønsker at tage ændringsforslag nr. 6 i betænkningen tilbage.
Ønsker nogen at optage dette ændringsforslag?
Da det ikke er tilfældet, er ændringsforslaget bortfaldet.

Hr. Søren Egge Rasmussen, Enhedslisten, værsgo.

Kl. 13:06

(Ordfører)
Søren Egge Rasmussen (EL):
Så kunne man høre, at der var et ændringsforslag, som røg ud af historien. Jeg synes, det er lidt tankevækkende, at vi for et døgns tid siden fik to ændringsforslag, som ændrer markant på det her lovforslag. Der er åbenbart blevet lavet en eller anden aftale i fredags, som har medført, at det bliver til en hovsalovgivning, og man har ændret lidt på mulighederne for ministeren til at give dispensation. Kommuner og boligforeninger pålægges, hvis den her lov, man er kommet frem til, vedtages, at reducere antallet af almene familieboliger til 40 pct. Nu lægges der så ind, at der, hvis man har nogle boligområder i de her hårde ghettoer, som nogle kalder dem, 16 områder, er en mulighed for, at man, hvis der er under 2.100 beboere, og hvis den såkaldte ghettolistes kriterier for meget kriminalitet ikke rammes, kan gå ind i en proces, hvor man ikke ender med at fjerne 60 pct. af de almene familieboliger. Så kan det være, man kan lave en dispensationsansøgning, så man ender på 50 eller 60 pct. almene familieboliger.

Jeg synes da, det er interessant, hvis det er den demokratiske proces, som har medført, at der alligevel er et flertal, som synes, at der var for mange uhensigtsmæssigheder i den såkaldte parallelsamfundspakke, som er på vej til at blive vedtaget. Men det fremgår ikke klart af ændringsforslagene, hvad der er historikken bag det. Hvad er det for nogle møder, der har været i fredags, hvem har deltaget, hvem skal have æren for de her to ændringsforslag? Det kunne da egentlig være interessant at få at vide, f.eks. her ved andenbehandlingen. Hvad er årsagen til, at man lægger det her snit lige ved de 2.100 beboere? Hvorfor er det ikke 2.134, for så kunne der måske være en boligforening mere, der kunne have gavn af det her ændringsforslag? Man kan også forestille sig mange andre muligheder i forbindelse med dispensation, for de her kriterier, der er lagt frem, er ikke rimelige, og det er jo det, som debatten har afspejlet – f.eks. det her kriterie med, at man ser på, hvad beboernes indkomst er, i forhold til hvad den er i gennemsnit i den region. Når man så har med Københavnsområdet at gøre, kommer man faktisk frem til, at hvis der bor nogen, som er på den høje dagpengesats, så tæller de sådan set med blandt de fattigste af beboerne. Er det tegn på, at de så udgør et problem? Det kan man ikke sige præcist. Så der er alle muligheder for, at man egentlig kunne finde på rimelige dispensationsbegrundelser for ikke at fjerne 60 pct. af de almene familieboliger.

Jeg synes, den her sag er dybt grotesk. Kunne man forestille sig, at der var et folketingsflertal, der ikke kunne lide parcelhuskvarterer – parcelhuskvarterer med familieboliger – og at et flertal i Folketinget så gik ud og sagde, at man nu ville fjerne 60 pct. af disse familieboliger, for man kan ikke lide beboerne derude, og så fandt man på en lovgivning, som kunne sikre, at det kom til at ske? Det ville være helt absurd, men det er sådan set det, der er ved at ske i det angreb på den almene sektor, som sker her ved anden og tredje behandling, som er planlagt i den her uge. Det er et angreb på hele den almene sektor, for med det, der bliver vedtaget her, ved man ikke, hvad det er for nogle områder, som bliver de næste områder, som kommer på de her lister, og som nogle vil kalde hårde ghettoer. Det er så absurd, og det løser ikke noget at pege fingre af boligområder og udskamme dem. Det, der skal til – for der er jo nogle steder, der er problemer, f.eks. med beskæftigelse – er jo en beskæftigelsesindsats, og den har været dybt fraværende i den proces, der har været for at komme frem til nogle aftaler mod parallelsamfund.

Boligordførerne fik en rapport i går, som gennemgår de her 16 områder, som man nu laver særlovgivning på, og jeg ville egentlig ønske, at vi kunne lave en god høring om de her områder. Det er boligområder, som er i en proces, hvor man nogle steder har brugt trecifrede millionbeløb på at renovere boligerne, og nu står man i den situation, at man har boligområder, som er toprenoverede, hvor folk

Figure 18: PDF example.