Speaking to voters: Explaining variations in the class-based support for social democratic parties through language complexity

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Philosophy in European Politics and Society at the University of Oxford.

Peter Sonneveld

Candidate number: 1040415

St Antony's College

Supervisor: Dr Matthias Dilling

Word count: 29,388

Michaelmas Term 2021

ABSTRACT

What explains variations in the class-based support for social democratic parties? Demand-size explanations have focused on the size and turnout of the working class, while supply-side explanations have focused on strategic policy maneuvering by parties. However, with both approaches, unexplained variation in the class-based support for social democratic parties remained. I introduce a novel argument, which I call the "language recognition thesis". The language used by politicians can be considered a way of symbolic representation used to appeal to voters. I test out this thesis in Great Britain, focusing specifically on whether and how language complexity in parliamentary debates influences class-based support for the social democratic Labour Party.

In preparing my variables, I draw on techniques from linguistics, sociology and computer science, using a linguistic formula and applying it to 1.8 million parliamentary speeches from the UK House of Commons, as well as distilling debate topics from these speeches using the machine learning technique Latent Dirichlet Allocation (LDA). After the data preparation, I use a two-step research strategy to investigate the proposed mechanism.

First, using a multilevel logistic regression, I investigate the relationship and its strength, through the use of the speech transcripts from the UK Parliament, which span a period of over thirty years, as well as thirty years of survey data.

Second, using a survey experiment, I attempt to provide evidence in support of my proposed mechanism, by subjecting participants to UK parliamentary speeches of varying complexity on a topic closely related to social class: welfare benefits.

The results show that when the language use across parties becomes more complex, people are more likely to rely on their class identity to determine what party they support. Working class individuals are more likely to support social democratic parties under these conditions, while members of higher classes are less likely to support these parties under the same conditions. Furthermore, people are more likely to support politicians who use less complex language. These findings provide important insights for the literature on class voting, social democratic party politics and political communications.

ACKNOWLEDGEMENTS

I have been here before, I thought as I first walked onto High Street in late September 2019, reminiscing about my earliest of Oxford memories - reading the book *Brideshead Revisited* by Evelyn Waugh and marveling in Charles Ryder's experiences with Sebastian and his extended family, and their time as students in the "city of aquatint" (Waugh, 2000: 17). After the book came the eponymous tv series with the hauntingly beautiful intro theme. This was what Oxford had always, from my childhood onwards, represented to me: a city of endless opportunities, a university of fascinating knowledge and people, a goal of a life in pursuit of knowledge. As I stood on the High Street that day, looking at the traffic passing Magdalen and the Examination Schools, I knew it: my dream had come true. I had arrived in the place where I had for so long felt like I should belong.

As I am writing this acknowledgement, having finished the two-year postgraduate course that brought me to Oxford, I find myself struggling to reconcile the actual experience I had in Oxford with the splendor of the Oxford image I grew up with. While the city is beautiful and the libraries are enormous, my unique "Oxford experience" of a pandemic-time Oxford combined with University-supported victimization by a fellow student has made me wonder: is Oxford really just a mirage?

We live in an era of shattered illusions and ongoing disappointment with institutional problems that lead to hurt among individuals. Unfortunately, Oxford is such an institution, with huge problems hurting the very individuals who, at first, are elated because of their acceptance to this bastion of knowledge and excellence, but, over time, have encountered the flip side of Oxford's legacy: reputation is everything. I suppose I was naive, thinking that the Oxford adage of "it's okay not to be okay" was true, only to find out that my feelings of insecurity and doubt fell onto deaf ears with the welfare team of my college, and led to victimization by a fellow student. Where safety and inclusivity is preached to the outside world, on the inside it's better to suffer in silence. To me, this has turned the city, and its renowned university, from an aquatint into a dim gray. In light of this, I feel there are very few people worth thanking in Oxford. That said, two are phenomenal exceptions.

First, my thesis supervisor, Matthias Dilling. From stimulating me to continuously improve my research (and especially the theoretical underpinnings) to giving me the freedom to study whatever I found most interesting (including allowing me to change my entire thesis topic midway through), his supervision was the shining light that pierced through the dark clouds

over Oxford. He showed himself to be that rare breed, not often seen in Oxford, of genius academics who still are empathetic human beings. My thesis would not have been what it is today had it not been for his considered and useful remarks on my work. I feel proud to have worked with him for two years.

A second person who needs to be mentioned here is Noah Bacine. We only met near the end of my time at Oxford, but I have found working with him to be both invaluable and thoroughly enjoyable. He introduced me to the wonders of experimental research, and kept pushing me to improve the design of my experiment. In addition to that, he was also a wonderful source of feedback on my ideas about experimental research as well as any questions I had during the process of creating the actual experiment (there were a lot). I feel grateful for his help and kindness. Working with him has made experimental research feel truly enjoyable.

Finally, my greatest gratitude goes out to my friends and family. It has become almost a cliché to say that you only really find out who your friends are in times of adversity - but it is true. I, for one, can happily say that I have true friends and family members, who were not quick to rush to judgment. Rather, they listened, supported and believed in me, even at times when I could not bring myself to do so. In paraphrasing Peter Pulzer, I would say that the love of friends and family is the basis in life, “all else is embellishment and detail” (Pulzer, 1967: 98). My friends and family continuously reminded me of who I truly am, and who I aspire to be. I owe this Oxford degree to their support and their love, and I will be forever grateful.

Table of Contents

Abstract 2
Acknowledgements 3
Table of Contents 5

Introduction 7
- Introduction 7
- Conceptualizing class-based support 8
- Explaining varying class-based support for social democratic parties 9
- The argument 11
- Outline of the thesis 13

Theory 16
- Two electorates 16
- Competing explanations 1: demand-side showing changing individuals 19
- Competing explanations 2: supply-side displaying class-mediating contexts 21
- My argument: language recognition thesis 24
- Conclusion and outlook 29

Methodology 1: discussing case selection, conceptualizing and operationalizing variables, and explaining modeling strategy 30
- Case selection 30
- Data sources 32
- Dependent variable: Labour support 34
- Contextual independent variables: Dalton's PI, language complexity and debate topic 35
- Individual-level independent variables: class and education 40
- Control variables: age, gender, ethnicity and Labour government 41
- Research strategy: two steps 42
- Conclusion and outlook 44

Results 1: contextual factors activate traditional class divisions in Labour support 45
- Descriptive statistics 45
- Base demographics model 46
- Bringing in the context: individual-level only 49
- Contextual models: introducing the contextual variables 54
- Conclusion and outlook 58

Methodology 2: explaining the experiment design 60
- Outline and sample selection 60
- Treatment 62
- Controls 62
- Analysis of test results using proportions test 63
- Conclusion and outlook 64

Results 2: experimental results show respondents more likely to support politicians using simpler language 65
- Description of the data and comparison with the general population 65
- Baseline differences conditioned by the treatment and control treatment 68

Differences per social class categories 70
Differences per education levels 72
Differences per partisan identification categories 74
Assessing influence of speaker party 75
Influence of speaker party per social class 77
Conclusion and outlook 80

Conclusion 82
Implications of the findings 83
Potential criticisms 85
Future work 86

References 88

Appendices A-F 95
Appendix A – Examples of parliamentary speeches with complexity scores 95
Appendix B – Debate topics with their 15 most frequent words 97
Appendix C – ANOVA Figures for the topic effects 100
Appendix D – Full version of Table 3 101
Appendix E – Pre-analysis Plan 103
Appendix F – Example experiment 108

Introduction

Introduction

"The opinion polls showed a marked swing of working-class sentiment away from the party", an analysis wrote about the support base of the UK Labour Party. It could have been taken from an analysis of present-day, with political scientists rushing to explain Labour Party performance in the 2021 Hartlepool by-election through a changing face of the working class (Abou-Chadi and Gingrich, 2021). However, the comment is much older, and refers to an era even before that; in the 1970s, Steven Beackon noticed a "shift of the Labour party away from its traditional base" since the late 1960s (Beackon, 1976: 231). It seems, therefore, that the traditional class-based support for the Labour Party has been a subject for discussion for decades.

What explains variations in the class-based support of social democratic parties? Explanations focusing on demand-side changes have emphasized societal changes leading to a declining size of the working class, as well as a move away from class emphasis and towards postmaterialist values. Meanwhile, explanations focusing on supply-side changes have emphasized the role of party behavior, noting that when parties converge ideologically, people will be disinclined to vote along class lines. However, both approaches leave unexplained variation of the class-based support for social democratic parties.

This thesis will contribute to the literature on this topic by introducing a novel argument on language use by politicians, more specifically the complexity of the language used. This argument will be embedded in the literature on symbolic representation (Pitkin, 1967), in order to explain how people feel "recognition" because of language complexity in politics, and why this matters for their political preferences.

I argue that the complexity of the language used in politics affects the individual-level characteristics that define support, such as social class and education-level. When the language used by politicians is *more* complex, it becomes *less* recognizable, forcing people to return to their traditional partisan identities. These, I argue, will refer to their class identity. When politicians use *less* complex language, the recognition increases, and people will be less inclined to support parties along class lines.

To evidence this argument against competing explanations, I will adopt a two-step research strategy focused on the case of Great Britain. In the first step, I use a multilevel logistic regression in order to test both my argument and the competing explanations. The findings support my argument: when the language in politics becomes more complex, people are more inclined to vote along class lines. This results in working-class people being *more* inclined to support Labour, and those who are middle class being *less* inclined to support Labour.
In the second step of my research, I use a survey experiment in order to more accurately measure the effects of language complexity on party support. I find that people are more likely to support politicians who use less complex language. Furthermore, I find that, by using less complex language, Labour politicians can increase their support from middle class constituents.

In the remainder of this chapter, I will conceptualize what I mean by class-based support. To do so, I will first review existing work on the relationship between social class and support for social democratic parties. Then, I will introduce the research puzzle. Finally, I end this chapter on an overview of the structure of the rest of this thesis.

Conceptualizing class-based support

"Class is the basis of British politics; all else is embellishment and detail," wrote Peter Pulzer once (1967: 98). At the time of writing, this was very much the case. Class differences defined party choice in Britain to a greater extent than anything else (Butler and Stokes, 1977: 67). And not just in Britain; in their seminal study of American voting behavior, Lazarsfeld and his colleagues concluded that "a person thinks, politically, as he is, socially" (Lazarsfeld et al., 2021: 27). Voting along class lines thus became nothing more than "the rational expression of individual differences in (primarily) economic interests that cluster as a result of shared characteristics deriving from similar conditions of employment" (Evans and Tilley, 2013: 113).

What this meant in practice was that those working in the upper echelons of a company (such as managers) were more inclined to consider themselves middle class, while those who worked in more "hands-on" types of jobs (such as welders, cleaners and, most prominently, factory workers) were more inclined to consider themselves working class (Butler and Stokes, 1977: 72). With class identity came party identity (Lipset and Rokkan, 1967). In Britain, among those identifying as middle class (and upper class), the Conservative party was the party of choice,

while those identifying as working class opted for Labour, a social democratic party (Butler and Stokes, 1977: 77).

This connection between a party and a particular social class did not come from nowhere. As soon as the working class was enfranchised, a political ideology representing their rights was formed. Socialist parties put workers' rights front and center in their policies, and, in doing so, hoped to attract majority support (Wickham-Jones, 1996: 13). Over time, however, it turned out that dreams of majority support for proletariat-only policies did not turn into reality (Esping-Andersen, 1985: 7). Therefore, in efforts to branch out to other electorates as well, social democracy emerged as an ideology for the "people" rather than just the "workers" (Esping-Andersen, 1985: 6-8). This strategy was quickly adopted by parties in countries like Denmark and Britain, where the Labour party became prominent after more emphasis was placed on "policies of a non-class nature" (Esping-Andersen, 1985: 8; Beackon, 1976: 237).

Despite this strategy of electoral broadening, the working class kept voting for social democratic parties. As one (former) British Labour-voter put it recently: "It stems back to your granddad, your dad, and it's passed on to you. You vote Labour because we're working class" (Clark, 2021). That said, the class-based support for social democratic parties did not always have the same strength. In the British case, it was noted that there was a "marked swing" of working-class constituents moving away from the social democratic Labour party in the 1960s (Beackon, 1976: 231). More broadly, voting according to class lines became less prevalent over time, even in countries like Britain where it once was prominent (Evans and Tilley, 2013).

Explaining varying class-based support for social democratic parties

Several explanations have been fronted for the decline in class-based support for social democratic parties. With regards to the British case, the move of the working class away from the social democratic Labour party has been attributed to changes in the nature of the working class. Goldthorpe and his colleagues, for instance, noted that increasing affluence of the working class led to a change in the *nature* of their support for Labour: "Whereas the traditional working class had a solidaristic attachment to the Labour party […], the new working class was more instrumental in its political orientation. Its members' support for Labour was largely conditional upon the performance of the party whilst in office" (Goldthorpe et al., 1968: 78; Beackon, 1976: 231). Where the "old" working class supported the social democratic Labour

no matter what, the "new" working class was less easily swayed and more prepared to critically assess Labour's performance (Beackon, 1976: 231). This, in turn, would mean that traditional class-based support was no longer a given for Labour.

Others noted an emergence of a "second" left-wing politics, more focused on middle-class voters and their interests (Manza et al., 1995: 145). Hindess advanced the idea of a "middle-class takeover", arguing that the middle class took the reins within the Labour party, forcing the party's positions to become less radical, eventually leading to "The Decline and Fall of Social Democracy" (Beackon, 1976: 231-2; Hindess, 1971: 170). After all, where Labour used to represent working class interests, a failure to continue to do so would mean the working class would have to look elsewhere for representation (Chamberlain and Moorhouse, 1974: 519-20).

Many have argued that the size of the working class has been in decline, causing the nature of class voting to change (Best, 2011: 284). After all, if there are no longer the traditional class cleavages, what remains of class voting (Clark and Lipset, 1991)? According to this view, political support became less class-defined, as did the support base for social democratic parties (Manza et al., 1995: 145). Others argued that intergenerational shifts moved society's main concerns away from "materialist" towards "postmaterialist" values, increasingly forcing out the relevance of the class cleavage (Inglehart and Rabier, 1986; Inglehart and Abramson, 1994; Inglehart, 1977; Abramson and Inglehart, 1992).

Meanwhile, others have argued that it was not a change in society that explains variation in class-based support for social democratic parties. Rather, it was party behavior itself that was causing the class-based support to decline. According to Beackon, a party that emphasizes "policies of a non-class nature" will most likely attract supporters who do not view the party "as a class party" (Beackon, 1976: 237). Moreover, others, advancing the *political choice thesis*, argued that the relevance of class voting "reflect[s] the outcomes of party behaviour" (Jansen et al., 2013: 47), and that "parties define the choices offered to voters and hence the active political cleavages within that society" (Tilley et al., 2021: 240).

Both these approaches, which can be grouped as representing a "demand-side" explanation (societal changes causing changed individual characteristics) and a "supply-side" explanation (party behavior), have not fully explained the observed variation in class-based support for social democratic parties. With regards to the demand-side explanations, findings of persistent

class differences despite the increased importance of postmaterialist values (Weakliem, 1991) have raised doubt about the explanatory power of these explanations. As for the supply-side arguments, these have been met with criticism of a limited explanatory power in certain time periods (Evans and Tilley, 2013: 112), as well as an observed lack of change in social democratic party support, despite the party behavior arguments assuming this would occur (Pulzer, 1987: 387; Heath, 2015: 181).

This means that there is a need for a new explanation to help us understand more about the variations in class-based support for social democratic parties. Because, crucially, the question remains open: what explains variations in the class-based support of social democratic parties? In response to this research question, I intend to contribute with a new answer.

The argument

In this thesis, I argue that this new answer is rooted in the language used by politicians. More specifically, the *complexity* of the language politicians use. I shall argue that when politicians use *less* complex language, they are signaling to voters an elite cue, which leads to an increase in *recognition* among voters. When voters recognize themselves in the politicians, they will feel more represented. As such, they will be more inclined to support these *less complex*-speaking politicians, as they can be considered accurate representatives of these voters.

In what follows, I shall distinguish between the following scenarios and actors when going through my argument, bearing in mind that I am investigating the argument in the case of Great Britain. First, *both* the social democratic Labour Party and the right-wing Conservative Party use *more* complex language[1]. Second, *both* Labour and Conservatives use *less* complex language. Third, Labour uses *less* and Conservatives *more* complex language. Fourth, Labour uses *more* and Conservatives *less* complex language. I shall go over these scenarios below.

In the first scenario, when the mean language use in politics (so, by all parties) becomes *more* complex, the recognition decreases. People will *feel* less represented by politicians who do not speak "their language". What they will then be forced to do, is return to what they already know in order to determine their vote: their class identity. This means that when the language used in

[1] I am assuming a two-party system here, consisting of Labour (social democratic) and Conservatives (right-wing).

politics becomes *more* complex, the traditional class-party association starts to re-emerge. As such, the working class will be *more* inclined to support social democratic parties, while higher-class voters will be *less* inclined to do so. The reason for working class individuals to return to supporting social democratic parties is that these parties not only have the historic relationship with the working class, but they also present policy offerings more in line with the working class' redistributive interests (Evans and Tilley, 2013: 89, 112).

In the second scenario, however, when *both* parties use *less* complex language, people will not be forced to return to their class identities, so I argue this scenario will *not* lead to increased working class support for Labour. For the third and fourth scenarios, a similar logic holds true. When one of the two parties uses *less* complex language, people will be *more* likely to support that party than the other party – irrespective of their class identity.

This is not a way of saying that working-class individuals are not equipped to understand more difficult language. Rather, using less complex language is, I argue, a way of bridging the "recognition gap" – the gap in "worth and cultural membership" between different groups in society (Lamont, 2018: 423). Using less complex political language can appeal to the group that displays the highest levels of political discontent. When less complex language is used, people will better recognize themselves in politics and will be more inclined to support politicians using that less complex language. As such, I term my argument the *language recognition thesis*.

In addition to the general mechanism, I argue that there are two more factors at play. First, the influence of topics under discussion. Various parties may have different "ownership" of issues, meaning they are the party most often thought of in relation to that issue (Budge, 2015). What I argue is that on a topic "owned" by a social democratic party (such as welfare state policy), the basic logic of the language recognition is reversed. Now, *more* complex language signals *more expertise*, which means that the voters will be inclined to accept this more complex language as a sign of expertise, rather than lack of recognizability. Therefore, on an "owned" topic like the welfare state, I argue that social democratic parties will benefit from more complex language.

Second, where *more* complex language is expected to revive traditional class-based support, I expect *less* complex language to be available as a tool for reaching out across class lines. In

other words, where *more* complex political language is expected to incentivize working-class voters to support social democratic parties, parties that use *less* complex language can "lure" these voters away, through the increased recognition displayed. Thus, a right-wing party, for instance, will be able to successfully appeal to a working-class constituency by adopting less complex language. This refers to scenarios 3 and 4, described above. In this thesis, I intend to prove the expectations outlined in this introduction.

Outline of the thesis

The rest of the thesis is organized as follows. In the next chapter, I will explain the theoretical foundations underlying my argument in more detail. I will explain more about the tension between two electorates for social democratic parties, as well as the explanations that have been offered through demand-side and supply-side theories. I introduce hypotheses for these theories, in order to test their explanatory strength. After this, I present the theoretical building blocks of my argument, embedding the argument in theories of representation and recognition. I explain in greater detail how my proposed mechanism works, and why I believe it to be a promising explanation for the puzzle of variations in the class-based support for social democratic parties. I introduce multiple hypotheses in order to test my assumptions.

In the third chapter, I present the majority of my research strategy. I explain why I am investigating the case of the UK Labour party, by outlining that it is an extreme case which provides a crucial test for my argument. If my argument has any explanatory power whatsoever, it should be visible in this case. After this, I present the data I am using, which I take from the *Parlspeech V2* dataset (Rauh and Schwalbach, 2020), the *British Election Study* and *British Social Attitudes* surveys (British Election Study, 2021a; UK Data Service, 2021), as well as the *Comparative Manifesto Project* dataset (Volkens et al., 2020).

Then, I explain how I operationalize my variables, starting with the dependent variable used throughout most of my analysis: *Labour support*. I then explain how I use Dalton's polarization index (Dalton, 2008) and the Flesch-Kincaid Grade Level score (Kincaid et al., 1975) to create testable scores for the independent variables of ideological polarization and language complexity, respectively. Then, I explain how I used a topic modelling technique called *Latent Dirichlet Allocation* (Blei et al., 2003) to determine different topics in the data, so I can assess the influence of topic expertise on the proposed mechanism, as described above.

I then move on to explaining my research strategy, which consists of two steps. The first step will be a multilevel logistic regression model, which I apply to a pooled dataset covering thirty years of British politics between 1989 and 2019, using the data from the sources described above.

In the fourth chapter, I move on to present the results of the regression analysis. After describing the specificities of the dataset, I first present baseline demographics models, in order to show how the individual-level characteristics impact people's propensity to support the Labour party. I find the literature arguing that social democratic parties move more towards higher-educated, middle-class voters (such as Kitschelt, 1994) to be correct. However, when I introduce my contextual variables, the story begins to change; when introducing language complexity into the model, I find traditional class and education cleavages to re-emerge. In addition to this, I find evidence supporting my assertion that the mechanism is reversed for the Labour-"owned" topic of the welfare state; more complex language on that topic has a large independent positive effect on Labour support. Overall, the regression analysis finds support for my argument that *more* complex language across parties leads to *more* class-based support for Labour.

In the fifth chapter, I present the design of the second step in my research strategy: a survey experiment conducted through the Centre for Experimental Social Sciences (CESS) at Nuffield College, Oxford. I explain my design and methodological choices, as well as the process of conducting the study. I show the two treatments that I use in the experiment, namely simple versus difficult language, and original versus altered language. The latter treatment is meant to control for the effect of altering original political speeches to a simplified or more complex form.

In the sixth chapter, I present the results of my experiment. After describing the demographic makeup of the respondent pool, I present the baseline effects, which confirm the language recognition thesis' base assumption: that people are *more* inclined to support politicians who use *less* complex language. Furthermore, I show that no effect of altering original speeches occurred. After this, I move on to measuring results in more detail, disaggregating voters along class, education and party identification lines. Finally, I confirm my assumption that politicians using simpler language can "lure" away voters across class lines, by showing that the Labour

party has a distinct advantage in using simpler language to attract not their traditional working class base, but rather the middle class voters.

In the final chapter, I bring my results together and offer a concrete analysis of what the results are telling us. I confirm that the language recognition thesis has potential explanatory power, at least in the case of the British Labour party. That said, I note that there are limitations to this study, such the limited scope of the experimental research. Finally, future work is suggested to look into investigating more topic differences, as well as investigating the relationship using more fine-grained survey data.

Theory: explaining Labour support through demand and supply, introducing the language recognition thesis

While the introductory chapter introduced the research puzzle of variations in the class-based support for social democratic parties, this chapter will present and develop the thesis' main argument to answer the research question of what explains variation in the class-based support for social democratic parties. The main argument posits that the complexity of the language used by politicians is an important factor influencing *under what conditions* social class matters for political support, and *to what extent*. This is because the language used can be viewed as an elite cue to the voter, signaling a "recognizable" message to the traditional constituency. As such, the "language recognition thesis" implies that variations in complexity in the language used by politicians acts as a contextual variable that activates support-defining individual-level characteristics.

The remainder of this chapter begins by diving into the literature on class voting in general and in the UK in particular, in order to identify this thesis' theoretical building blocks. It then moves on to the competing explanations that have been fronted in the literature. These are divided in two categories: demand-side and supply-side explanations. Testable hypotheses are formulated in order to test these theories' veracity. After this, the language recognition thesis is presented. The theoretical expectations derived from this are then turned into testable hypotheses that will be assessed in later chapters in order to put the argument to the test.

Two electorates

From the very beginning of working-class enfranchisement, a political ideology that represented workers' rights was formed. Inspired by Marx' comment that it was the working class' "great duty" to conquer political power, socialist political parties were formed by those who, as Przeworski put it, "decided to utilize political rights of workers in those societies where workers had them and to struggle for these rights where they were still to be won" (Przeworski, 1985: 7-8). The decision to enter politics, made possible by the expansion of suffrage to lower classes, was made because of an interest not to engage in high-level politicking or form alliances with other parties, but rather to use politics as a forum for propaganda and organization (Przeworski, 1985: 9). After all, the revolutionary spirit had not died, and some considered a workers' revolution more useful than entrance into politics. However, entering elections was

also seen as a way of showing the country's leadership the "revolutionary fervor of the masses" (Przeworski, 1985: 10).

In the end, political abstention or merely symbolic participation were never real options for these parties, as their goal of advancing workers' rights could be met by competing in democracy and winning progress, even just in small steps (Przeworski, 1985: 10-11). Participation in the electoral arena was necessary to advance the proletarian agenda, and, so the socialists believed, soon the masses would vote to empower this political ideology. The parties put workers' rights at the center of their politics, and in doing so hoped to attract majority support (Wickham-Jones, 1996: 13). As Przeworski put it, socialists at the time believed that "[h]istory spoke through the people, people spoke in elections, and no one doubted that history would make people express their will for socialism" (Przeworski, 1985: 18). Soon, this idea seemed to be confirmed, through large victories in elections all over Europe (Przeworski, 1985: 18-19). Until 1918, the working-class parties saw themselves strictly as class parties (Esping-Andersen, 1985: 7).

However, while there was a clear electoral gap that was filled by socialism, the strict proletarian majority, it eventually appeared, remained out of reach (Esping-Andersen, 1985: 8). What emerged was a dilemma. Majority support for an ideologically pure workers party seemed out of reach, but abandoning the parties' class images would risk losing their very *raison d'être* (Esping-Andersen, 1985: 8). With the proletariat not becoming a numerical majority, and hopes that dissatisfied former middle-class members would "either become proletarians or join the army of the unemployed" failing to come true, a different strategy was needed (Przeworski, 1985: 23). The dilemma came down to this: either stay pure and never achieve your goals, or dilute your class image and attempt to achieve (modified versions) of your goals (Esping-Andersen, 1985: 8; Przeworski, 1985: 24).

It became clear that a transformation of socialism was necessary to continue to advance workers' goals. Social democracy therefore became an ideology of the "people" rather than just of the "workers", an ideology more attuned to modern society, and more willing to participate in parliamentary democracy (Esping-Andersen, 1985: 6-8). By engaging in politics, the very thing that traditional proletariat parties had refused to do, the social democrats escaped "political isolation" (Esping-Andersen, 1985: 9). While they still represented the same core constituency of workers, the main difference was that for the old parties, "program dictates

parliamentary power", whereas for social democratic parties "parliament dictates program" (Esping-Andersen, 1985: 8). In that sense, social democracy became a more practical version of the same class struggle, a version which allowed others to join in as well, and which was willing to put up with the realities of politics in order to achieve its ultimate goals.

To achieve these goals, however, being less stringent in their ideology was not enough. Knowing that the core constituency of workers would not achieve a numerical majority, the goal of gaining the power to enact socialist policies would only be met if an appeal would be made to voters from other classes (Przeworski, 1985: 24). This electoral pragmatism was quickly adopted in countries like Britain and Denmark (Esping-Andersen, 1985: 8). In Britain, the Labour party rose to prominence after including "policies of a non-class nature" to their party platform ever since 1922 (Beackon, 1976: 237).

Over time, the broadening of the social democratic appeal led to a second constituency. Contrary to the traditional working-class constituents, these were more highly educated, more progressive middle-class voters (Kitschelt, 1994: 286). This created a dilemma for social democratic parties. They wanted to continue to appeal to their socialist, conservative core constituents, but this at times could conflict with the new, more progressive constituency. Whereas the core constituency preferred traditional, economically left-wing yet culturally conservative views (Lefkofridi et al., 2014; Houtman, 2003; Lipset, 1959: 485; Przeworski, 1985), the new constituencies were higher educated, middle-class voters who preferred culturally liberal and economically moderate policies (Kitschelt, 1994: 33; Gingrich, 2017: 50).

This tension culminated most prominently in the 1990s. Some authors suggested social democratic parties ought to seek out "an electoral constituency built around sophisticated industrial technicians and engineers, white collar employees and middle managers, and the large sector of professionals in personal services" (Kitschelt, 1994: 301). This would help the parties to attract "younger women whose disposition toward libertarian values and center-left economic policies is stronger than that of any other group" (Kitschelt, 1994: 301). As such, the appeal to the traditional core seemed to have become less important, especially because many scholars noted that the relevance of the class cleavage on election outcomes had declined (Kitschelt, 1994; Esping-Andersen, 1985; Przeworski, 1985; Gingrich and Häusermann, 2015; Oesch, 2008a; Rennwald and Evans, 2014; Clark and Lipset, 1991; Best, 2011: 284).

Competing explanations 1: demand-side showing changing individuals

An influential branch of the literature has suggested that changes in society increased the prominence, in the social democratic support base, of the higher-educated, middle-class electorate over the traditional working-class constituency. One such change that has been discussed is the overall decline in the size of the working class (Best, 2011: 284). In Great Britain, for instance, the proportion of the working class in the electorate declined from roughly 60% in 1964 to about half that in 2010 (Heath, 2015: 180-1). At the same time, the size of the middle class increased, and since the mid-1980s, the middle class has become the largest occupational group in the country, comprising over 50% of the population in 2010 (Heath, 2015: 180-1). This, therefore suggested a society wherein class became heavily realigned (Manza et al., 1995: 145).

Others, most prominently Inglehart, argued that changing intergenerational attitudes moved interests of potential support bases away from "materialist" to "postmaterialist" values, like the environment or equality (Inglehart and Rabier, 1986; Inglehart and Abramson, 1994). Because the younger generations are more interested in these postmaterialist values, as time went by (through generational replacement), these postmaterialist values were becoming increasingly more important than the material ones, demanding parties to reposition themselves in relation to these issues (Abramson and Inglehart, 1992; Manza et al., 1995: 146).

Because of these societal changes, the social democratic parties have been incentivized to strategically branch out to this newer part of the electorate. After all, whether the size of the working class has declined or whether value change means voter attitudes change, sticking with the traditional core constituency and presenting workers' rights no longer seemed to be a winning strategy – making up for the lost votes can only be done by changing what group needed to be appealed to (Kitschelt, 1994: 286). Thus, social democratic parties since the 1980s started to gradually appeal more to the middle-class constituents, and less to the traditional working class (Dalton, 2021). As a result, some argued, the middle class started voting for the left, while the traditional working class moved to the right-wing parties (Weakliem, 1991; Oesch, 2008b: 349-350).

If these theories are right, we should expect to see that the working class no longer supports social democratic parties, at least when looking at support since the late 1980s or early 1990s.

Instead, support for social democratic parties should be higher among the middle class, as they are the group that the social democratic parties, with their now culturally liberal, economically moderate policies, have vied for in this time period. We should expect to see that, out of the two electorates available to the social democratic parties, the newer, middle-class electorate should be the one most attracted to these types of parties. To see if this is correct, I introduce the first hypothesis on this matter.

Hypothesis 1 *– Working class constituents since 1990*[2] *are less likely to support social democratic parties than middle-class constituents.*

One other change in society that is often denoted is the change in levels of education among citizens. Over time, it is argued, citizens have become more and more educated. In European countries, the number of individuals receiving a university-level degree has, as a share of the population, doubled between 1992 and 2018 (Ford and Jennings, 2020: 300-1). Similar increases in educational expansion have been found in the United States (Inglehart, 1977: 293). As a result of the increase in education, more and more citizens, it was argued, became equipped to make political decisions independent of their class or other social loyalties (Manza et al., 1995: 144; Jansen et al., 2013: 48). Values changed as a result of the educational expansion, as higher educated citizens were found to be more libertarian, focusing more on individual freedom, and less interested in formal hierarchies (Kitschelt, 1994; Weakliem, 2002). According to Inglehart (1977: 321), the "political skills" were rising among the public, as a result of this educational expansion. Because of this increase in the voter skillset, through the process of "cognitive mobilization", these voters could "participate in politics at a higher level" (Inglehart, 1977: 321). As a result, the higher-educated voters have tended to prefer political parties that focus on liberal, culturally progressive, and internationalist values (Ford and Jennings, 2020: 301).

Given that the social democratic parties have appealed to this higher-educated, middle-class group, through culturally progressive and economically centrist policies (Kitschelt, 1994: 33; Abou-Chadi and Wagner, 2020; Dalton, 2021), the lower-educated voters have often been found to move to other parties, such as the extreme right (Arzheimer, 2009: 268). If, therefore, the influence of education is relevant, and the societal changes in terms of educational

[2] This aligns with the centric moves by several social democratic parties in Europe (Dalton, 2021), as well as the time period I am looking at in my thesis, which is 1989 to 2019 (see chapter 3 for more on this).

expansion affect support-levels for social democratic parties, we should expect to see that lower-educated individuals are less likely to support social democratic parties than those who have more years of formal education under their belt. To see if this holds, I come to my second hypothesis.

Hypothesis 2 *– Lower-educated constituents since 1990 are less likely to support social democratic parties than higher-educated constituents.*

Overall, what these individual-level explanations imply is that the traditional support base for social democratic parties, the working-class, lower-educated individuals, have been substituted in terms of relevance by the middle-class, higher-educated newer constituency.

However, while individual-level factors as explanations for varying levels of class-based support have been frequent and prominent in the literature, they have not always been able to account for the full story. Even the very calculation on which the decline in class-based voting was based, the Alford Index, has attracted a great deal of criticism, due to its oversimplification of class categorization (Manza et al., 1995: 147-8; Evans, 2000: 407-8). Some other calculations of class sizes have found class voting to remain influential in explaining political support (Van der Waal et al., 2007: 416).

Furthermore, while Inglehart's postmaterialism thesis has attracted much attention, Weakliem found that, despite value changes, class differences across cohorts remain intact (Weakliem, 1991: 1354). Moreover, some authors have found that lower-educated voters are not less inclined to vote for social democratic parties when these parties propose progressive policies (Abou-Chadi and Wagner, 2020: 259; Manza et al., 1995: 150). Therefore, looking at the individual-level characteristics seems to be unable to tell the whole story. This warrants an investigation of the supply side of politics.

Competing explanations 2: supply-side displaying class-mediating contexts

While the demand-side explanations focus most on societal changes at the individual level (in the sense that class realignment and educational expansion change individual voter attitudes), supply-side explanations have been better at contextualizing the changes in the class-based support for social democratic parties. According to these explanations, the relevance of class

voting "reflect[s] the outcomes of party behavior" (Jansen et al., 2013: 47) and "parties define the choices offered to voters and hence the political cleavages within that society" (Tilley et al., 2021: 240). This approach is not exclusively top-down, in the sense that it does not think changing shapes of class groups or education have no effect; rather, explanations in this realm assume that, for instance left-wing parties that are office-seeking, are "expected to place less emphasis on traditional working-class issues" as fewer manual laborers can give less electoral rewards to parties vying for their vote (Jansen et al., 2013: 50).

The emphasis on supply-side explanations is not new. In the 1970s, Beackon argued that the British Labour party had emphasized "policies of a non-class nature" since 1922, and therefore would attract only those who would not view the party in line with the traditional social democratic class identity (Beackon, 1976: 237). However, by appealing to groups beyond their traditional class core identity and de-emphasizing class, the nature of political conflict, some argued, could no longer be centered around class conflict, weakening the class cleavage's influence on politics (Manza et al., 1995: 146; Przeworski, 1980: 43). Given the "positive connection" between Labour's policies and those attracted to the party (Beackon, 1976: 237), it has been argued that party behavior by Labour and other social democratic parties could increase or decrease class-based voting.

Most prominent in explaining the supply of party politics in order to explain class voting has been the *political choice thesis*, which argues that "patterns in class voting reflect the outcomes of party behaviour" (Jansen et al., 2013: 47, 72). In this view, it is not so much that party behavior explains all voting behavior. Rather, class voting is "conditioned by the extent to which political parties are seen to be associated with the interests of different social classes" (Jansen et al. 2013: 50). In this view, class identities are not irrelevant. No, class identities and their relevance to the vote are *conditioned* by party maneuvering. If a party positions itself as a working-class party, more working-class support can be achieved, and if a social democratic party appeals more to the middle class, their working-class support is likely to diminish (Jansen et al., 2013: 50).

As Evans and Tilley (2013: 88-9) argue, ideological policy offerings matter a great deal here. With the decline of party identification, increasing education levels and the rise in media consumption, voters became more responsive to party's performance and policy offerings (Evans and Tilley, 2013: 88). Evans and Tilley say that voters with different resource-levels at

their disposal are inclined to have different preferences with regards to redistributive policies (Evans and Tilley, 2013: 89). Put simply, the "have nots" will want redistributive policies, while the "haves" will not be inclined to support parties that propose redistribution.

The problem, however, is that parties no longer ideologically differ as much anymore. Since the 1980s, especially social democratic parties have moved towards the center, in their attempts to appeal to the middle class (Dalton, 2021). Furthermore, as Beackon argued, social democratic parties like Labour have been reluctant to even emphasize a class identity since the 1920s (Beackon, 1976: 237). What happens is that class remains important in predicting a person's ideology, but ideology no longer predicts party choice – according to Evans and Tilley, this is because of a lack of ideological distinctiveness between parties (2013: 112).

What the authors proposing the *political choice thesis* show, is that when the ideological differences between political parties are smaller, there is a corresponding decline in the relevance of class voting (Jansen et al., 2013: 72). Both ideological convergence between traditional class-representing parties in Britain (Evans and Tilley, 2013: 112) and a lack of ideological divergence across different countries (Jansen et al., 2013: 72) have been shown to affect levels of class voting. In that sense, ideological differences at the party level *influence the effect of* individual characteristics. When parties offer more "class-related political choices", people will be more inclined to vote along class lines (Jansen et al., 2013: 72).

What we would therefore expect to see is that if there is *more* ideological polarization in a party system (meaning parties differ more ideologically), there will be *more* class voting. And because social democratic parties have been one of the parties "historically representing class interests" (Evans and Tilley, 2013: 112), we should be able to see more class-based voting for social democratic parties occur *in this context* of ideological polarization. Hence, I formulate my third hypothesis.

Hypothesis 3 *– More ideological polarization between parties leads to more class-based support for social democratic parties.*

Contrary to what the demand-side explanations presented above imply, the supply-side story shows that the traditional support base for social democratic parties does not *need* to be lost.

The political choice thesis argues cogently that class-based support for social democratic parties can be reinvigorated by ideological distinctiveness.

The political choice thesis, successful as it has been at explaining much about varying levels of class voting, has not been without criticism. Even the authors proposing the explanation show that they see problems. Jansen and colleagues, for instance, note that, using their coding choices, they cannot replicate the findings of a "collapse of class voting" in the British 1997 elections, even though Evans and Tilley, in the same book, argued that this did occur (Jansen et al., 2013: 73). Evans and Tilley themselves point out that they have been unable to explain the decline in class voting between the 1960s and mid-1970s (2013: 112).

Furthermore, others have argued that changes in policy offerings by social democratic parties did not negatively affect their electoral fortunes. Pulzer, for instance, showed that most social-democratic parties across Europe in the 1980s performed similar to how they did in earlier times, despite the tension of the two potential electorates (working class and middle class) that the parties had to appeal to (Pulzer, 1987: 387). Furthermore, while ideologies between Labour and Conservatives in Britain drifted apart in the 1980s, there was no corresponding increase in class voting – even though this would be expected if the political choice thesis offered a sufficient explanation (Heath, 2015: 181). This shows that the political choice thesis does not explain the full picture.

My argument: language recognition thesis

According to Heath (2015), the gaps left by the political support thesis can be solved by looking at representation, using Hanna Pitkin's definition (1967). Where policy offerings are a form of "substantive" representation in the sense that a party may substantively represent a certain class when its policy offerings align with the policy interests of this class, he argues that "descriptive" representation matters too, as it occurs when party representatives "mirror" the social characteristics of those the party aims to represent (Heath, 2015: 175). So, in other words, a party representing working-class constituents should have working-class representatives in order to effectively represent their constituents.

The potential for descriptive representation was shown through experiments in America and Britain, where results showed that working-class voters considered working-class political

candidates as people who understood what problems these voters encountered – an effect that was stronger than those of "gender, education, race, experience, or even political party" (Carnes and Lupu, 2016: 839-9). It is not so much that the working-class candidates actually always pursue working-class policies, but in the mind of the voters, they will most likely do so (Heath, 2015: 176). Therefore, descriptive representation seems to serve as exactly the kind of elite cue voters like to use to give them a sense of rationality when making their vote choice (Bawn et al., 2012: 577).

The problem with descriptive representation is that it may have potential explanatory power, but in practice, politicians are almost always "wealthier, more educated, and more likely to come from white-collar jobs than the citizens who elect them" (Carnes and Lupu, 2016: 832). In Britain, while a sizeable group of working-class citizens remains, the proportion of working-class MPs has declined steadily over time, to only a small percentage today (Carnes and Lupu, 2016: 836; Heath, 2015: 182). At the same time, the relationship between Labour and the working class did not correspondingly weaken as the number of working-class MPs declined (Heath, 2015: 180-2).

What is needed, then, is a framework that explains why people are inclined or disinclined to vote along class lines, in a way that deals with what is actually happening in politics. What is the context that mediates a person's propensity to vote along class lines? This is where my argument comes in. Where Pitkin notes that representation means "to make present something that is *not* in fact present" (1967: 92, italics in original), what becomes clear is that working-class representation in politics does not require working-class individuals to actually be *in* politics. Rather, the working class (and all other classes, for that matter) need to be *made present* by those who are in politics.

As Pitkin notes, the "crucial test" for political representation is whether the representative is *believed in* by the represented (1967: 102). Those politicians who successfully are *believed in* by the public become accepted as the people's "symbolic representative" (Pitkin, 1967: 102). When a politician wants to represent a voter, they need to act in a "recognizable" manner. Whether they share social characteristics with the voter or not, whether they offer class-related policies or not, they need to *communicate* their representative status to those they represent. I argue this is done through their *language*. More specifically, the *complexity* of the language they use in their communications.

I propose that the complexity of the language politicians use can serve as an elite cue, signaling to voters that they are "standing for" them. Using the right complexity of language will bridge any "recognition gap" between the politician and the voter, as the voter will feel like their "language", their cultural membership, will be made present in politics (Lamont, 2018: 423). Voters will *recognize* the language used by the politician, as it will be the language they themselves use when interacting with their peers. As such, I term my argument the "language recognition thesis".

There is preliminary evidence supporting this theoretical reasoning. Less complicated language in party manifestos in Germany and Austria meant that party positions were better communicated to the voters, which resulted in voters being better at placing the parties on a left-right scale (Bischof and Senninger, 2018: 474). Moreover, populist parties already tend to "simplify" policy challenges in their manifestos and use simpler language, which is part of their anti-establishment appeal (Lin and Osnabrügge, 2018: 2, 7; Decadri and Boussalis, 2020; Marx and Schumacher, 2018). Furthermore, Spirling (2016) found that, in Britain, after the Second Reform Act was introduced (which doubled the size of the electorate), cabinet members were more likely to use less complicated language in parliamentary debates, hoping to appeal to the new electors.

How language complexity connects to class voting is as follows. I argue that when a politician uses *less complex* language, they are directly appealing to voters who prefer less complex language. This bridges the recognition gap, for instance with voters who prefer politicians who "[s]peak normal" (Noordzij et al., 2020: 8). When politicians, however, do not actively signal to these voters through their language (meaning they continue to use more complex language), these voters will need to explore different avenues of determining their vote. This, I argue, will lead to an increase in class-based support for social democratic parties. If a voter feels no *recognition* in politics (because they cannot relate to it), determining their vote choice becomes a matter of returning to what they already know: voting along class lines[3].

[3] Or not vote at all, as was found to be the case for working class and "less educated" individuals in Britain in the 2000s (Evans and Tilley, 2017b: 711). As I investigate party *support* rather than *vote* (see chapter 3 for more on my outcome variable), I do not further investigate the lower turnout rates among working class and less educated individuals in this thesis.

What does this mean in practice? As language complexity mediates the strength of individual characteristics defining partisan support, I argue that when the language across parties becomes *more* complex, working-class voters will be *more* inclined to support social democratic parties, while higher-class voters will be *less* inclined to do so. As such, a class divide will occur. I assume a two-party scenario in the case I am investigating (Great Britain), with Labour as the social democratic party, and the Conservatives as their main rival. The scenarios I envision therefore become that *both* parties use *more* complex language (scenario 1), that *both* parties use *less* complex language (scenario 2), that Labour uses *less* and Conservatives *more* complex language (scenario 3), and that Labour uses *more* and Conservatives *less* complex language (scenario 4). The scenario where *both* parties use *more* complex language (scenario 1) will lead to *more* working class support, and *less* higher-class support for Labour. However, when *both* parties use *less* complex language, I expect this will *not* lead to more working class support for Labour. When only one party uses less complex language (scenarios 3 and 4), people are more likely to support that party, whatever their class identity may be.

Noting the strong association between social class and education (Weakliem, 2002: 145), I further argue that when the language in politics becomes *more* complex, lower-educated voters will be *more* inclined to support social democratic parties, while higher-educated voters will be *less* inclined to do so. The reason these groups go back to supporting social democratic parties, is that these parties have both the historic relationship with these social groups and the policy offerings that align with their redistributive interests (Evans and Tilley, 2013: 89, 112).

For the avoidance of doubt, I am in no way suggesting that working-class and lower-educated voters are unable to understand more complicated language. Rather, my suggestion is tied to the higher levels of political discontent found among lower-educated, working-class individuals (Noorzij et al., 2020: 2). In a study performed in the Netherlands, this group tended to mistrust politicians who "beat about the bush" and use "academic language" (Noordzij et al., 2020: 8). The individuals feel a cultural distance to the political class, a lack of recognition, and an important factor in this is the communications' style used by politicians. As one interviewee in that study noted, "[w]henever you ask a question […] to which they could just answer with a yes or no […] then you get an endless reply [meaning that] you lose track of the story" (Noordzij et al., 2020: 8). Others in the study corroborate this, indicating that they would prefer politicians to use language used by "common" people, which is identified as "straightforward, without

limits", "upfront" and using "normal words [and] not much official language" (Noordzij et al., 2020: 8).

Considering all of the above, I formulate two hypotheses in order to test my argument. In the below, when I write "across parties", I mean the language used by *all* parties. In other words, the hypotheses refer to scenario 1, as described above.

Hypothesis 4 *– More complex language across parties leads to more class-based support for social democratic parties.*

Hypothesis 5 *– More complex language across parties leads to more education-based support for social democratic parties.*

In addition to this, I feel a crucial nuance is warranted. While language complexity in general can affect class voting, the manner in which it does so can be mediated by the *topic* that is under discussion. It could be that different topics allow for politicians to more (or less) effectively use language complexity to appeal to the electorate. After all, if a politician or their party is believed to be "issue owner" of a certain topic (Budge, 2015), their "believability" might be higher when they actually use *more complex* language rather than less. The logic is simple: when someone uses more complex language on a topic they "own", they may be seen as an expert on that topic. Thus, by using more complex language on a topic they are known for, they are "believed in" (Pitkin, 1967: 102), at the very least more than those who are not the "owners" of that topic.

With regards to social democratic parties, one such "owned" topic is the *welfare state*. Social democratic parties have a history of using the state apparatus to enhance the welfare of the citizens (Przeworski, 1980: 52). Whether people agree with them or not, social democratic parties have "ownership" on the topic of the welfare state. I therefore expect these parties to benefit when discussions on this topic become more complex. In order to test this expectation, I formulate another hypothesis. Again, with "across parties" I am referring to the *mean complexity of all parties* that participate in these debates.

Hypothesis 6 *– More complex language across parties in debates on the welfare state leads to more support for social democratic parties.*

One final thing that is worth discussing is the party effect. As argued above, I argue that when the language in politics is *more complex*, people return to their traditional class-tied party. But what if *both* parties use *less* complex language (scenario 2) or language complexity *differs* across parties (scenarios 3 and 4)? I expect this to take away the class (or education) tie to partisan support. But in order to find support for this expectation, I need to add one final test. And in doing so, I turn to the class-party relationship that is at the center of this thesis: the one between the working class and social democracy. Can the traditional support base of the social democratic parties, the working class, be appealed to by other parties if they use *less* complex language? In order to test this, I formulate my final hypothesis:

***Hypothesis 7** – By using less complex language, other parties can successfully appeal to the traditional support base of social democratic parties – the working class.*

Conclusion and outlook

In this chapter, I started out by explaining the historic tension for social democratic parties, needing to balance two electorates that require different approaches. After this, I discussed several demand-side and supply-side explanations that have been offered to explain the varying class base of social democratic parties. I outlined their strengths and supporting evidence, and pointed out the deficiencies that still require answering. After that, I introduced my argument, the *language recognition thesis*, and outlined the underlying logic supporting the argument: that varying levels of language complexity used in politics can affect the recognizability felt by voters, in turn influencing the strength of class voting. I introduced several hypotheses to test out my expectations.

In the next chapters, I will outline my research strategy and perform my analysis, in order to assess these hypotheses. As my research strategy is two-step, the next chapter will introduce, in addition to the data used as the basis for both study approaches, the methodology for my multilevel logistic regression model. The chapter directly following will then show my regression results and analysis. In the chapters after that, the methodology for my experiment will be discussed, and the experimental results will be analyzed.

Methodology 1: discussing case selection, conceptualizing and operationalizing variables, and explaining modeling strategy

Having discussed the theoretical framework for this thesis, it is now time to explain the means of testing the language recognition thesis and answering the research question that is driving this thesis. To do this, I shall first explain the case selection and the rationale behind it. After that, I will introduce the key variables that will be looked at in the analysis. This includes both the conceptualization and operationalization of the variables, and connecting both to the existing and original data I am using for measurement. After that, I will explain my two-step research strategy. This chapter will explain the modelling strategy used. The experimental design is explained in a later chapter.

Case selection

My analysis involves an extensive single-case study of British politics, with a specific focus on the British Labour party. This is for two reasons.

First, the United Kingdom is an example of an *extreme* case (here I follow the *extreme* case typology offered by Seawright and Gerring, 2008). This is because of the extreme value on the explanatory value of interest: class voting. As Seawright and Gerring argue, "it is the rareness of the value that makes" such a case valuable (2008: 301). The choice for a single-case study of the United Kingdom as an *extreme* case "serves as an entrée into the subject" (Seawright and Gerring, 2008: 302) in order to display its potential explanatory power, after which further scrutiny using for instance cross-case analysis can be performed. Significant results in this case will encourage future, cross-national comparisons using the observed relationship between language complexity and the class-based support for social democracy.

After all, the class-party linkage had historically been particularly strong and prominent in the UK (Evans and Tilley, 2013: 87; Pulzer, 1967: 98, Butler and Stokes, 1977: 67). However, over time, this cleavage has declined to a point where the directors of the British Election Studies (BES) felt compelled to conclude that, at the end of the twentieth century, social class "had come to play a very limited role in determining the voting preferences of the British electorate" (Clarke et al., 2004: 50). That said, research since has shown that elite cues and party behavior can change the influence of class on voting behavior (Evans and Tilley, 2013; Jansen et al.,

2013; Evans and Tilley, 2017a; Robison et al., 2021; Tilley et al., 2021: 240). If there would be any country where the *political choice thesis* would need to be true if it had any value, it would be in the United Kingdom (as shown in Evans and Tilley, 2013).

It follows that the same goes for my hypothesized *language recognition thesis*: if the hypothesized relationship between language complexity and the class-based support for social democratic parties will be true anywhere, then it should be true in the United Kingdom, with its rich history of the class cleavage (Lipset and Rokkan, 1967). This makes it a promising initial test for my novel argument. Finding support for my argument would encourage investing resources into a more comprehensive cross-country analysis. In turn, finding no support for my argument would indicate that it is even less likely to find support in other contexts.

Critics may argue that choosing a case that would actually be *least* likely to confirm (or, indeed, most likely to *dis*confirm) the proposed relationship would be a stronger strategy. After all, if the mechanism would be confirmed in such a case, the mechanism's viability could be considered very strong. However, choosing a least likely case and finding null results would perhaps disconfirm the mechanism *in that case*, but not in *all* cases. Conversely, choosing a case where the hypothesized relationship *should* be present *if* my argument had any explanatory power, as I do here, helps make clear the potential for a cross-case relationship. If the relationship is confirmed here, future studies should look into its viability across a wider population of cases. If the mechanism is not even confirmed in this most likely case, there might not be reason to investigate further among new cases.

The second reason is that the election system in the United Kingdom creates a virtual duopoly between the Conservatives and Labour – and has done so for many decades (Lijphart, 2012: 13-4). Because of this, the United Kingdom provides an excellent testing ground for the impact of language complexity used by politicians, and the link between social democracy and the working class. There are few countries left in (western) Europe where the social democratic party is still such a large and influential party within a virtual two-party system (one other clear example is Spain[4]). What makes selecting the UK even more useful, is the relative absence of

[4] I chose the UK over Spain as a case to investigate, because Spain only became democratic again in 1977, making it a new democracy for the early stages of my observed time period (Orriols, 2013: 361). The UK, meanwhile, has seen stable democracy, including class cleavages, for much longer. Furthermore, on a practical note, I do not speak Spanish, which would have complicated analysis of the language used in politics as well as analysis of relevant sources.

populist (radical right) parties in the national parliament, which means that there is no rhetoric of large populist parties in the language used in politics. This lack of populist parties in the system facilitates the conditions for the first rigorous test of my novel argument. By investigating the UK as a case, I can investigate a relatively stable party system, with the main parties staying the same (Labour and Conservatives), for the entire duration of the time period I am looking at.

Data sources

In order to test the hypotheses (see previous chapter), I turn to several data sources. With regards to the dependent variable of support for the Labour party and the demographic variables (see below), I turn to two data sources that are specific to Great Britain: the British Election Study (BES) and the British Social Attitudes (BSA) surveys (British Election Study, 2021a; UK Data Service, 2021). While the BES is conducted after general elections, the BSA is run every year. While the BES in itself contains all the information needed for the analysis, additional data is needed to allow for yearly data analysis in this study, which is why I include the BSA data. More specifically, I use the BES surveys from 1992, 1997, 2001, 2005, 2010, 2015, 2017 and 2019 (British Election Study, 2021a). I combine these with the BSA annual surveys from 1989 to 2019, with the only missing year being 1992 (UK Data Service, 2021). With regards to the BSA, I would have liked to use the year 1992 as well, but in that year, the BSA was not performed, as money was being reallocated to the BES post-election survey of 1992 (Brook et al., 1994: 1). In order to use the data for a long-term model, I brought together the data from the BES and BSA surveys on the variables of interest, giving me a pooled dataset with yearly values from 1989 to 2019.

In order to test out the third hypothesis on ideological polarization between parties, I have used data from the *Comparative Manifesto Project* (Volkens et al., 2020), using the *Manifesto Corpus* (version 2020-2; Burst et al., 2020) through the *ManifestoR* package (Lewandowski et al., 2020). In line with the rest of the data, I used data from the same time period. This means that I used manifesto data for the British general elections of 1992, 1997, 2001, 2005, 2010, 2015, 2017 and 2019 (Burst et al., 2020).

Finally, in order to test out the hypotheses on the influence of the complexity of the language used by politicians, I have chosen to use parliamentary speeches made in the House of

Commons as a data. While many studies focusing on a party's message tend to look at party manifestos as a data source (see for instance Crabtree et al., 2020), I believe that parliamentary speeches provide rich and meaningful data for my analysis. This is supported by a growing body of literature investigating the use of parliamentary speeches as data.
One obvious benefit of using parliamentary speeches over manifesto texts is that the former is a much more fine-grained data source. While manifesto texts only appear once per election cycle, parliamentary speeches appear on an almost daily basis throughout election cycles, which means this type of data allows for more nuanced over-time comparison (Valentim and Widmann, 2021: 7; Kosmidis et al., 2019: 818-9).

Moreover, parliamentary speeches have been shown to attract a lot of media attention (Valentim and Widmann, 2021; Bäck et al., 2019: 184; Marcinkiewicz and Stegmaier, 2019: 577; Proksch and Slapin, 2012). Furthermore, parliamentary speeches are used by both parties (Bäck et al., 2019: 165) and individual parliamentarians (Baumann et al., 2015; Benoit et al., 2012; Lauderdale and Herzog, 2016: 392; Poyet and Raunio, 2020: 20) to communicate with voters. Finally, parliamentary speeches have been shown to have a strong effect on electoral outcomes (Marcinkiewicz and Stegmaier, 2019: 577, 592), something which may be known to parliamentarians themselves, as the content of these speeches were found to change when elections are imminent, suggesting strategic use of parliamentary speeches (Proksch and Slapin, 2012: 521).

As the data source for the parliamentary speeches used in this thesis, I have taken the UK House of Commons portion of the *Parlspeech V2* dataset (Rauh and Schwalbach, 2020), a dataset which consists of parliamentary speeches from parliaments in nine countries[5], for a period of up to 32 years. From the House of Commons, the *Parlspeech V2* dataset contains 1,956,223 speeches, spanning a period of roughly 31 years. All speeches are transcribed verbatim. The speeches are of varying length, with most speeches not exceeding a few hundred words. In order to exclude simple interruptions, points of order, or passing remarks, I only include speeches of 50 tokens or more in my analysis. I also exclude speeches made by chairpersons (the Speaker), as these generally only speak to give the floor to a Member of Parliament (MP) or to structure a debate. This approach gives me 1,817,395 speeches from the House of Commons, delivered between 22 November 1988 and 17 December 2019.

[5] Austria, the Czech Republic, Germany, Denmark, the Netherlands, New Zealand, Spain, Sweden, and the United Kingdom.

Dependent variable: Labour support

For my dependent variable, I follow the approach suggested by Evans and Tilley (2013: 92-3), by choosing party support rather than vote choice as my dependent variable. However, unlike Evans and Tilley, I restrict my party support variable to a dichotomous variable of whether someone supports the Labour Party or not: *Labour support*.

The reasons for choosing *support* over *vote choice* are threefold. First, it allows for the meaningful operationalization of party preference even in non-election years (Evans and Tilley, 2013: 93). Taking *vote choice* as the dependent variable would not allow for this. Second, *support* better explains the long-term links between the Labour Party and the social classes, as *support* is less influenced by short-term effects of campaign influences (Evans and Tilley, 2013: 93). Using *support* rather than *vote choice* therefore should better display the changes in the class-based support for the Labour Party. Third, opting for *support* rather than *vote choice* helps eliminate the non-random error introduced to vote choice measures by tactical/strategic voting that is common in majoritarian systems like the British election system (Evans and Tilley, 2013: 93; Lijphart, 2012: 14-5). In other words, while looking at who voted Labour would likely include those who are not generally supportive of Labour, but preferred them in a certain election over other large parties, the *Labour support* variable should accurately display who is (and who is not) a supporter of the Labour party.

I derive the *Labour support* variable from the pooled dataset of the BES and BSA surveys I created. I use the same variables that Evans and Tilley (2013: 93) use. From the BES surveys, this means using the following question:

> *"Generally speaking, do you think of yourself as Conservative, Labour, Liberal Democrat or what?"*[6]

From the BSA surveys, I used the results of the following three questions, noting that every participant (eventually) is asked about their partisan preference[7]:

[6] The options "Nationalist/Scottish National" and 'Plaid Cymru' were added for respondents in Scotland and Wales, respectively.

[7] Together, these questions act as filter questions before someone is asked to disclose what party they support – if any (see for instance British Social Attitudes, 2014: 16).

"*Generally speaking, do you think of yourself as a supporter of any one political party?"*

[IF NO] "Do you think of yourself as a little closer to one political party than to others?"

[IF NO] "If there were a general election tomorrow, which political party do you think you would be most likely to support?"

Evans and Tilley show that taking the results from the three BSA questions together "gives a very similar number of non-identifiers to the single BES question" (2013: 93). This means that, using this approach, a yearly party support variable can be derived.

As my interest is primarily in the Labour Party, I turn the results of this combined variable from the BES and BSA surveys into a dichotomous variable, with "1" indicating that someone is a Labour supporter and "0" indicating that they are not. I deal with missing data as follows. All cells that are either empty or contain "NA" or "NaN" (meaning "Not a Number") are coded with a '0'. The same goes for the answer choice "Don't know". The reason for this is that *if* people are indeed Labour supporters, they are likely to have filled out Labour as their answer. It is unlikely that people who identify themselves as Labour supporter would not fill out "Labour" as the answer to this question. The only exception to this is the answer option "Refuse to answer" (or variations like "Prefer not to say"). As the respondents giving this answer may very well be Labour supporter, but ones who are unwilling to divulge their partisan identity in a survey, these respondents cannot be ruled out as Labour supporters (but of course also not definitively confirmed). For this reason, I exclude these respondents from the analysis. As this group is very small[8], this is not expected to lead to problems in the analysis.

Contextual independent variables: Dalton's PI, language complexity and debate topic

In order to test the hypothesis on ideological polarization effecting the class-based support for Labour, I require a method of determining parties' ideological positions across different elections, and the ideological difference between the parties as a result of those positions. To

[8] In the complete dataset of 132,929 rows, a total of 367 respondents gave an answer like 'Refuse' to the party support question and are therefore excluded. This comes down to 0.276% of the entire dataset.

do so, I take the party manifesto positions as measured by the Comparative Manifesto Project (CMP) in 'rile', their left-right scale (Volkens et al., 2020; Laver and Budge, 1992: 24-5). This variable's scoring is calculated by subtracting the "sum of percentages" a party scores on left-leaning issues by the party's scores on right-leaning issue categories (Bakker and Hobolt, 2013: 32-3). This measure was investigated by Bakker and Hobolt (2013: 41), who found that the traditional left-right measure correlates highly with both an adjusted version of the CMP measure (0.91) and with two expert surveys (0.90). This gives me confidence that the party's left-right positions are accurately reflected in this measure.

That said, my main interest when investigating the first hypothesis is not on the ideological positions *per party*, but *across all parties* – after all, I am merely interested in levels of ideological polarization. This means that, similar to Jansen et al. (2013: 55-6), I am interested in a measure that takes a weighted mean left-right position across all parties. Luckily, such a measure exists: Dalton's Polarization Index (Dalton, 2008). The measure is calculated using the standard deviation from the mean party position in a party system, weighted by party size (Jansen et al., 2013: 55-6). The index is written as follows:

$$PI = \sqrt{\sum (party\ vote\ share_i) * (\frac{party\ L/R\ score_i - party\ system\ average\ L/R\ score}{5})^2}$$

where *i* represents individual parties (Dalton, 2008: 906). The index has a value of 0 when no polarization takes place (meaning all parties have the same positions), and 10 when polarization is at the extreme (meaning all parties are at either end of the left-right scale) (Dalton, 2008: 906).

To create a variable called *Dalton PI* for my analysis, I apply this formula to the CMP data for the UK for the years 1992, 1997, 2001, 2005, 2010, 2015, 2017 and 2019. Others have noted that one problem with this measure is that the CMP only includes "significant" parties into its datasets, which could mean that not all parties are covered in the CMP data (Jansen et al., 2013: 55). However, as this thesis is focused on UK politics, which functions in a highly disproportionate system that favors the (two) main parties (Lijphart, 2012: 14), the potential of missing a relevant party in the analysis is limited. That said, the CMP data contains only 3 UK parties for the 2010 and 2015 elections (Volkens et al., 2020), which means potentially relevant

parties are left out of the equation. Unfortunately, there is no easy remedy to this when using CMP data.

For the rest of the hypotheses, I use the second independent variable, *language complexity*. I measure this variable using a measure commonly used to measure language complexity in (political) texts: the Flesh-Kincaid Grade Level score (Kincaid et al., 1975: 14). This measure is an adjusted version of the traditional Flesch Reading Ease (FRE) score, which was developed to determine the "readability" of texts on a range from 100 (easy for any literate person to read) to 0 (practically unreadable), and in doing so establish the difficulty of texts (Flesch, 1948: 229). The FRE then needs to be converted to figure out the number of years of education a reader needs to understand the level of complexity (Douma, 1960: 23). The Flesch-Kincaid Grade Level formula, however, already adjusts the FRE to give a grade level score, thereby removing any potential conversion error (Kincaid et al., 1975: 19).

Furthermore, the Flesch-Kincaid measure is better accustomed to mature audiences and text contents, as it was created to calculate the education level Navy personnel needed in order to understand their training material (Kincaid et al., 1975). A further benefit of the Flesch-Kincaid measure is the scaling used. Jansen (2011: 497) points to the difficulty in assessing how much of an improvement a FRE score of 50 is compared to a score of 40. The Flesch-Kincaid score remedies this, as it directly displays the number of years of education needed to understand the text (Jansen, 2011: 497). The Flesch-Kincaid Grade Level score is calculated using the following formula (Kincaid et al., 1975: 14):

$$GL = 0.39 \times \left(\frac{words}{sentence}\right) + 11.8 \times \left(\frac{syllables}{word}\right) - 15.59$$

The outcome of the formula shows the grade level required to understand the text. Therefore, higher Flesch-Kincaid scores correspond to higher complexity, which is different from the FRE. Higher complexity in this context means either *longer words* (more syllables per word) or *longer sentences* (more words per sentence), or *both* (Schoonvelde et al., 2019: 5).

The FRE and Flesch-Kincaid measures have been used in a range of different articles and books, with studies varying from the text complexity of political science research (Cann et al., 2014), the communication of monetary policy (Jansen, 2011), the complexity adjustments German parliamentarians implement in their speeches when addressing lower-educated constituents

(Lin and Osnabrügge, 2018: 3), the language complexity used in debates and campaign speeches by American presidential candidates (Wang and Liu, 2018: 303), and the impact of the Second Reform Act on linguistic complexity of speeches by British MPs (Spirling, 2016: 121), among others (for an overview, see Schoonvelde et al., 2019: 5-6). Admittedly, these measures are not perfect, and Benoit, Munger and Spirling (2019) have introduced a crowd-coded measure that has a stronger theoretical basis for analysis of political text. That said, in the same article, the FRE is found to correctly predict 72% of the human coders' judgments on the most difficult text among two text snippets (Schoonvelde et al., 2019: 6; Benoit et al., 2019: 502). This good performance, combined with the financial constraints to my thesis research, and the common usage of the Flesch(-Kincaid) measure, means I am confident that using the measure will produce accurate scores for my analysis.

The reason I choose to use the Flesch-Kincaid measure rather than the FRE is because it automatically gives a clear and understandable score (years of education needed to understand the text), and that it is better suited for texts for mature audiences. I apply the Flesch-Kincaid formula described above to the 1.8 million parliamentary speeches taken from the *Parlspeech V2* dataset, to get over-time scores of the complexity of the language used in Parliament. The mean value of the speeches in the dataset is 14.411126, with a standard deviation of 5.214608. Examples of parliamentary speeches with varying Flesch-Kincaid scores assigned can be found in Appendix A.

As my dependent variable (*Labour support*) is measured in yearly instances (through the used BES and BSA surveys), I am, unfortunately, forced to take yearly means of these language scores. That said, I believe there is still enough variation to have a meaningful analysis, as this approach results in 31 different *language complexity* scores[9]. For the sake of comparison: the Dalton polarization index relies on manifesto data, and therefore only has 8 different score values across the dataset[10].

In order to measure the last hypothesis, I include the third independent variable: *debate topic*. The reasoning behind this is that parliamentary speeches do not take place in a vacuum; rather,

[9] I did investigate whether using the bi-annual Eurobarometer data would be a way of finding more fine-grained data, but the questions in that survey were not as consistent over time as in the BES and BSA surveys.
[10] For the years in my dataset that go between election years, I use the polarization index score for the last election year, as this is the most recent manifesto data for such a year. So, for instance, I use the polarization index score for the election year 1992 for the non-election years 1993, 1994, 1995 and 1996.

each speech takes place within a debate on a certain topic, for instance the national budget or health care policy. Several studies have been conducted into the different debate topics (or themes) in parliamentary speeches (over time). Greene and Cross (2017), for instance, found that European Parliament plenary debate topics vary greatly over time, and the incidence of debate topics is influenced by exogenous events such as the Euro Crisis or referenda on European treaties. Curran et al. (2018) found that debate topics in the New Zealand parliament varied over time. Furthermore, they found that certain parties were more prone than others to emphasize certain topics – for instance, the right-leaning National Party started debating economic topics more after 2008 (when they gained government power) at the expense of an emphasis on law and order, whereas the Labour party saw no such change (Curran et al., 2018: 8).

In order to make sure the debate topics in my UK dataset have significantly different complexity tied to them, I decided to perform an ANOVA test. First, however, I needed to determine the different topics, as the *Parlspeech V2* dataset used as the basis for this thesis does not contain debate topics (Rauh and Schwalbach, 2020). Therefore, I turned to *Latent Dirichlet Allocation* (LDA), the most commonly used technique for topic modelling (Blei et al., 2003; Blei, 2012; Curran et al., 2018: 3). LDA considers all documents in the collection under analysis (here: parliamentary speeches) to share the same set of topics, but every document contains the different topics to varying proportions (Blei, 2012: 79). As such, a topic can be considered a distribution of keywords, which occur in different proportions across the different texts (Curran et al., 2018: 3).

As such, LDA is an example of a *probabilistic* topic modelling algorithm; LDA helps understand the probability of a topic occurring in a text (Blei, 2012: 79). A crucial step when performing topic modelling analysis is the selection of k, which refers to the number of topics (Curran et al., 2018: 3). There are no set rules for this, and the number of topics can actually vary over time (Greene and Cross, 2017: 87). Key is not to select too few topics, as this would group different topics together, but also not too many, as this would lead to the same broad topic being displayed twice (for instance, a separate topic for ‘roads’ and one for ‘trains’, rather than the overarching ‘infrastructure’ topic).

For this analysis, I selected $k = 20$ topics as a mean topic number over the 31-year period of the analysis. This is based on the number of ministerial departments listed on the UK Government

website (UK Government, 2021a). This list includes 23 departments, but they include the offices of the Leader of the House of Commons and the Leader of the House of Lords, as well as 'UK Export Finance', which is a part of the Department for International Trade (UK Government, 2021a). These three are therefore excluded, bringing the total to 20 topics. After applying the LDA model on the 1.8 million parliamentary speeches in my database, I ended up with 20 debate topics (see Appendix B for the topics and the 15 most frequent words per topic).

In order to check the difference in complexity, I needed to find out the difference in means between the complexity scores for the various topics. Therefore, I performed an ANOVA test. I tested the influence of *topic*, which ($p<0.05$) showed a significant mean difference. As a further check, I checked the two assumptions of ANOVA (homogeneity of variances across groups and normally distributed data). Appendix C displays the plots for these, which confirm that the assumptions of homogeneity of variances and normality are not violated. As such, I feel confident in assuming there is significant difference in mean topic complexity score, which is why I can use the topic complexity in my analysis. In line with what has been discussed in the theory chapter, I focus here on one topic: the welfare state.

As can be seen, the main independent variables in my analysis so far are contextual variables, at the system level. This impacts the modelling strategy, which will be described later on in this chapter. Below, I will first describe individual-level demographic variables that I use as *individual-level* independent variables: social class and education.

Individual-level independent variables: class and education

The main individual-level variable, crucial to determine the strength of the class-based support for the Labour Party, is *social class*. There are two main ways of determining a respondent's social class that are used most frequently. There is the option of asking a respondent what social class they think they belong to (if any), an approach that was for instance taken by Robison and colleagues in their study of class-based appeals (Robison et al., 2021: 732). I will also use this approach in the experiment that is described later on in this thesis.

The other approach is looking at a respondent's occupation. This is the approach taken by Jansen and colleagues in their cross-national study of class voting (Jansen et al., 2013: 53). Furthermore, it is the approach that is strongly preferred by the authors of the BSA for studies

investigating "how the attitudes of different social classes have changed over time" (NatCen Social Research, 2017: 19). The BSA authors include a recoding guide, in order to turn the different measures of class that have been used over time into one comparable measure (NatCen Social Research, 2017: 19). Following their guide, and applying this to the BES data as well, I end up with a five-category class scheme: *Salariat (Higher & Lower)*, *Clerical (Junior non-manual)*, *Petty Bourgeois*, *Foremen/Technicians*, and *Working class*. I perform listwise deletion on missing data.

A second independent variable is *education level*. Education is a social characteristic that is "likely to be correlated with both party preference and social class" (Evans and Tilley, 2013: 94). Others have also found education and social class to have a strong association (Weakliem, 2002: 145). Including this variable allows me to test both hypotheses 2 and 5. To measure education level, I have opted to use the variables asking respondents what their highest qualification is, a measure that appears in both the BES and BSA surveys. From this, I have recoded the results into a six-category classification: *No qualifications*, *Below GCSE*, *GCSE*, *A-level*, *Undergraduate* and *Postgraduate*. I used listwise deletion to filter out the missing data. I also removed answers like "Don't know" and "Refuse to answer", as well as those who had foreign qualifications.

Control variables: age, gender, ethnicity and Labour government

I also added control variables based on both respondent's demographics and context. These include, at the individual level, the respondent's age (in years), gender (1 for male, 0 for female) and ethnicity. With regards to age, I also add a log transformed version of the age variable, to account for "the rate of change per year [being] inversely proportional to age" (Karon, 1964: 27). For ethnicity, I collapsed the answers into the five-category classification proposed by the UK government: "*White"*, *"Mixed or Multiple Ethnic Groups"*, *"Asian or Asian British"*, "*Black, African, Caribbean or Black British"*, and *"Other ethnic group"* (UK Government, 2021b). I did not include *religion* as a variable, as it has been found to be of lesser relevance to British politics (Evans and Tilley, 2013: 113).

With regards to the contextual-level control variable, I include a dichotomous variable *Labour government* to denote when the Labour Party was in government (1 meaning yes, 0 meaning no). This is to account for potential influence of government status on voter support. Some, like

Heath and colleagues noted that people tend to be turned away from the ruling party, especially in second-order elections (Heath et al., 1999: 394). That said, others find that there is a potential electoral bonus for incumbents (Eggers and Spirling, 2017: 908). Including this control variable therefore allows me to account for government status effects.

Research strategy: two steps

I will investigate the research question and the hypotheses using a two-step research design, consisting of a modelling approach and an experiment (which is described in chapter 5). As the first stage, in order to investigate hypotheses 1 through 6, I will estimate a multilevel logit model with individual variables at the first level, and contextual variables at the second level. This approach is similar to the work of Arzheimer and Carter on political opportunity structures for right-wing extremist party success (2006: 428). Through the estimation of this model, I can view the support for the Labour party based on the individual respondent's social class and education level in itself and *within the context* of ideological polarization, language complexity, topic complexity and Labour government status (the contextual variables). I will use the survey waves of the BES and BSA used for clustering. The model is described below.

Following the notation for a combined multilevel logit model suggested by Hox and colleagues (Hox et al., 2018), the individual-level model looks like so:

$$L_{is} = \beta_{0s} + \beta_{1s}X_{1i} + \beta_{2s}X_{2i} + \beta_{3s}X_{3i} + \beta_{4s}X_{4i} + \beta_{5s}X_{5i} + \beta_{6s}X_{6i} + e_{is}$$

with L_{is} referring to the individual respondent's (i) support for the Labour party in cluster s (the survey wave), X_{1i} referring to the respondent's social class, X_{2i} to their education level, X_{3i} to their age, X_{4i} to the log age, X_{5i} to their gender, and X_{6i} to their ethnicity, with e_{is} referring to the error term.

The formulas for the explanatory variables at the contextual level are, for the intercept

$$\beta_{0s} = \gamma_{00} + \gamma_{01}P_s + \gamma_{02}C_s + \gamma_{03}D_s + \gamma_{04}G_s + u_{0s}$$

and for the slopes

$$\begin{aligned}
\beta_{1s} &= \gamma_{10} + \gamma_{11}P_s + \gamma_{12}C_s + \gamma_{13}D_s + \gamma_{14}G_s + u_{1s} \\
\beta_{2s} &= \gamma_{20} + \gamma_{21}P_s + \gamma_{22}C_s + \gamma_{23}D_s + \gamma_{24}G_s + u_{2s} \\
\beta_{3s} &= \gamma_{30} + \gamma_{31}P_s + \gamma_{32}C_s + \gamma_{33}D_s + \gamma_{34}G_s + u_{3s} \\
\beta_{4s} &= \gamma_{40} + \gamma_{41}P_s + \gamma_{42}C_s + \gamma_{43}D_s + \gamma_{44}G_s + u_{4s} \\
\beta_{5s} &= \gamma_{50} + \gamma_{51}P_s + \gamma_{52}C_s + \gamma_{53}D_s + \gamma_{54}G_s + u_{5s}
\end{aligned}$$

$$\beta_{6s} = \gamma_{60} + \gamma_{61}P_s + \gamma_{62}C_s + \gamma_{63}D_s + \gamma_{64}G_s + u_{6s}$$

with P_s referring to the *party system polarization* for survey *s*, C_s referring to the *mean language complexity* for survey *s*, D_s referring to the *mean language complexity per topic* for survey *s*, and G_s referring to whether there was a Labour government in survey *s*. The *u*-terms refer to the (random) residual error terms for survey *s* (Hox et al., 2018: 9). The formula for the intercept predicts the average *Labour support* in survey *s* by the party system polarization (P), language complexity (C), topic complexity (D), and Labour government status (G). In other words, if γ_{01} is positive, it means that people are, across all surveys, more likely to support Labour if there is more party system polarization – and vice versa (Hox et al., 2018: 9). The same goes for the other contextual variables: complexity, topic complexity, and Labour government status.

The slope formulas refer to the relationship between the independent variable in question and the outcome variable within the context of the contextual variable. In other words, $\gamma_{12}C_t$ refers to the relationship between social class and Labour support *dependent on* the language complexity in political speeches. If γ_{12} is positive, this means that the effect of social class on Labour support is higher when the language used in Parliament is higher. Therefore, the language complexity can be seen as a *moderator variable*; the strength of the relationship between the individual-level predictor variable and the outcome variable (Labour support) varies according to the value of the moderator variable (Hox et al., 2018: 9). Furthermore, as the effects of the demographic variables may vary across different survey waves, I also estimate random slopes for the demographic variables in the full model.

In order to view the model in one equation, with six individual-level variables and four contextual-level variables, the model can, through substitution and rearranging, be written as one complex regression equation like so:

$$\begin{aligned} L_{is} = {} & \gamma_{00} + \gamma_{10}X_{1i} + \gamma_{20}X_{2i} + \gamma_{30}X_{3i} + \gamma_{40}X_{4i} + \gamma_{50}X_{5i} + \gamma_{60}X_{6i} + \gamma_{01}P_s + \gamma_{11}X_{1i}P_s \\ & + \gamma_{21}X_{2i}P_s + \gamma_{31}X_{3i}P_s + \gamma_{41}X_{4i}P_s + \gamma_{51}X_{5i}P_s + \gamma_{61}X_{6i}P_s + \gamma_{02}C_s + \gamma_{12}X_{1i}C_s \\ & + \gamma_{22}X_{2i}C_s + \gamma_{32}X_{3i}C_s + \gamma_{42}X_{4i}C_s + \gamma_{52}X_{5i}C_s + \gamma_{62}X_{6i}C_s + \gamma_{03}D_s \\ & + \gamma_{13}X_{1i}D_s + \gamma_{23}X_{2i}D_s + \gamma_{33}X_{3i}D_s + \gamma_{43}X_{4i}D_s + \gamma_{53}X_{5i}D_s + \gamma_{63}X_{6i}D_s \\ & + \gamma_{04}G_s + \gamma_{14}X_{1i}G_s + \gamma_{24}X_{2i}G_s + \gamma_{34}X_{3i}G_s + \gamma_{44}X_{4i}G_s + \gamma_{54}X_{5i}G_s \\ & + \gamma_{64}X_{6i}G_s + u_{1s}X_{1i} + u_{2s}X_{2i} + u_{3s}X_{3i} + u_{4s}X_{4i} + u_{5s}X_{5i} + u_{6s}X_{6i} + e_{is} \end{aligned}$$

It is worth noting that, as can be seen in the model above, the random error terms u_s and others are connected to the X_i. This is because the explanatory, individual level variable X_i and the

corresponding error term u_s are multiplied, resulting in an error term that is different for different values of X_i, the situation referred to as heteroskedasticity in ordinary multiple regression analysis (Hox et al., 2018: 10). Therefore, using a multilevel model is more sensible to use than an ordinary multiple regression in a situation like this.

Conclusion and outlook

Having first established the rationale for my case selection and discussing my data sources, in this chapter I conceptualized and operationalized the variables I will use in my analysis. After this, I described the first stage of my research design. The second step will follow in chapter 5.

In the next chapter, I will discuss the results of the regression exercise in detail. The chapters after that will contain discussion of the experiment and the results thereof.

Results 1: contextual factors activate traditional class divisions in Labour support

Having established the theoretical foundations of my argument, and having outlined my methods for data preparation and testing, it is now time to present the results of my multilevel logistic regression model. After first discussing the descriptive statistics in the dataset, I move on to the testing of my hypotheses. In order to test hypotheses 1 and 2, as formulated in the theory chapter, I run baseline logit models including only the demographic variables (social class, education level, age, log age, gender, and ethnicity), so I can determine the effects of these individual characteristics on support for the Labour Party, independent of context. After this, I place the individual-level characteristics into a contextual model, in order to determine the effects of party system polarization (hypothesis 3), language complexity (hypotheses 4 and 5) and welfare state complexity (hypothesis 6). After this, I bring all the findings together and formulate preliminary takeaways before it is time to move on to the second study in this thesis, the experiment.

Descriptive statistics

To learn more about the data I am using in my regression analysis, I first present some descriptive statistics. It must be noted, however, that a substantial amount of the variables is either binary (dichotomous) or unordered categorical. This is the case for the individual-level variables *Labour support*, *gender* and *Labour government* (dichotomous variables), *social class, education level* and *ethnicity* (categorical variables). Including these into a descriptive statistics table would be either impossible (for the categorical variables) or would not add much information (for the dichotomous variables). Therefore, these are not included in the descriptive statistics, which can be viewed in Table 1[11].

Descriptive statistics for numerical variables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Dalton's PI	77605	1.158	0.638	0.729	0.773	1.104	2.694
Language complexity	77605	14.424	0.193	13.982	14.314	14.483	14.831
Welfare state complexity	77605	12.252	0.418	11.447	11.991	12.438	13.188
Respondent age	77605	49.943	18.012	18	35	64	99
Respondent log age	77605	3.840	0.388	2.833	3.555	4.159	4.595

[11]All tables and figures in this thesis were created using version 5.2.2 of the *R* package *stargazer* (Hlavac, 2018) and version 2.8.9 of the *R* package *sjPlot* (Lüdecke, 2021).

When looking at these statistics, it first becomes clear that the general level of polarization in UK politics seems to have been low during the observed time period; after all, the scale of the polarization index ranges from 0 to 10 (Dalton, 2008). Therefore, the observed mean of 1.158 and standard deviation of 0.638 show that, by and large, polarization over the 31-year time period is limited.

With regards to the complexity scores, remember that the complexity score used represents the number of years of education needed to understand the text (Kincaid et al., 1975). A one-unit increase in complexity score would therefore mean that a reader would need one more year of education to be able to understand the text. The general complexity seems to have limited variation, with a mean of 14.424 and standard deviation of just 0.193 This is likely due to the limited data points, as a result of the imbalance between the original text data (daily) and the survey data used (yearly). Combining this data requires taking yearly means from the text data's complexity scores, resulting in the lower variation[12].
What can be further seen in the descriptive statistics is that the topic complexity for a topic "owned" by Labour, the welfare state (see Appendix B for the topics), has a lower mean complexity (12.252) and a higher standard deviation (0.418) than the overall complexity. This, in addition to the ANOVA results discussed in the previous chapter, confirm an effect of debate topic on language complexity, and justify including this variable in the model.
Finally, the age variable ranges from 18 to 99, with the mean respondent aged (almost) 50.

Base demographics model

To start analyzing the data, I first present the results of the individual-level variables on the outcome *Labour support*. This is to show the effects of these variables without the contextual variables. Doing so allows me to answer hypotheses 1 and 2. Table 2 displays the results of the demographics models, by first presenting the demographic variables in separate models, before combining them into one demographic model.

As can be seen, there are class-based effects on a person's likelihood of supporting Labour in both the class-only model and the full demographic model – except for working class. This is

[12] For comparison, the original text dataset shows a mean language complexity score of 14.411, with a standard deviation of 5.215.

as predicted, and in line with other literature claiming that Labour "lost" the working class vote in the 1990s and has not (yet) recovered it (Evans and Tilley, 2013: 112). Given that my pooled dataset covers exactly the time period (1989-2019) that includes the introduction of New Labour and its move away from the traditional working class base, it makes sense that the data presented here confirm that when a person is working class, this does not in itself have a significant effect on their likelihood of supporting Labour. Thus, presented with the results here, hypothesis 1 can be confirmed.

When we investigate the effects of education, a different picture emerges. All education categories have highly significant positive effects on a person's likelihood of supporting Labour, in both the education-only model and the full model. Education, therefore, significantly influences whether or not someone is likely to support Labour. That said, while all education categories are significant, some are more significant than others. Especially university-educated individuals are much more likely to support Labour. It seems that more education makes one more likely to support Labour. This is unsurprising, as it is in line with the social-democratic move-to-the-middle that Kitschelt predicted in the mid-1990s; appealing to the higher educated, the younger women who favored libertarian values and center-left economic policies (1994: 301). The effects of New Labour seem to be present here as well.

The prediction that the least educated group, the ones with "no qualifications", would be less likely than higher educated groups to support Labour holds true. While still positive, the effect is much smaller than among those with university degrees. What this, and the positive effects among the other "lower" educated categories, indicate is that people, across different education levels, are still positively inclined to support Labour – they are just not *as* positively inclined as higher educated individuals. In other words, the traditional Labour base is not necessarily *disinclined* to support Labour, they are just not the group Labour seems to be targeting anymore. As such, the effect among lower educated individuals remains smaller than among those who are higher educated. Hypothesis 2 can be confirmed.

Other noteworthy findings, among the control variables, are that there seems to be a small negative (yet highly significant) effect of someone's age on their likelihood to support Labour in the age-only model (odds ratio of 0.99). That said, as age invariably changes over time, the build-up of this small effect can, over time, lead to a stronger negative effect. In the full model,

Table 2: Baseline Demographic Model

Dependent variable: Labour support

	Social class only	**Education only**	**Age only**	**Gender only**	**Ethnicity only**	**Combined model**
Predictors	*Odds Ratios*	*Odds Ratios*	*Odds Ratios*	*Odds Ratios*	*Odds Ratios*	*Odds Ratios*
(Intercept)	0.06 ***	0.06 ***	0.04 ***	0.07 ***	0.15 ***	0.11 ***
Foremen or Technicians	1.70 ***					1.84 ***
Petty Bourgeois	2.22 ***					2.43 ***
Salariat (Higher & Lower)	1.26 ***					1.23 ***
Working class	0.95					1.03
Below GCSE		1.41 ***				1.39 ***
GCSE		1.19 ***				1.18 ***
No qualifications		1.20 ***				1.23 ***
Postgraduate		3.26 ***				3.35 ***
Undergraduate		1.80 ***				1.82 ***
Age			0.99 **			1.00
Log Age			1.30			0.92
Male				1.09 ***		0.92 ***
Asian or Asian British					0.84	0.93
Black, African, Caribbean or Black British					0.75 *	0.92
Mixed or Multiple ethnic groups					0.91	1.01
White					0.49 ***	0.60 ***
Observations	77605	77605	77605	77605	77605	77605
R^2 Tjur	0.006	0.006	0.000	0.000	0.002	0.015

** p<0.1 ** p<0.05 *** p<0.01*

Note: reference categories are Clerical (junior non-manual) for social class, A-levels for education, 18 years for age, female for gender, and Other ethnic (including Arab) for ethnic group.

however, age does not show an effect change. Moreover, the additional variable of the log transformed age does not have a significant effect in either model.

With regards to gender, a striking difference appears. When looking only at gender, men are (slightly) more likely to support Labour, but when all demographic characteristics are combined, this effect moves towards a small negative (both significant effects). What this means is that men are, in general and across all other categories, more likely to support Labour, but the minute other factors, such as class, education or ethnicity come into play, this effect changes.

Ethnicity has no significant effect, *except* for the "White" group. In fact, across models, the "White" category is the one with the strongest negative effect on a person's propensity to support Labour. This is an interesting finding, yet not an unsurprising one. Studies have found Labour to be more appealing to non-white voters, and Conservative to be more appealing to whites (Skinner et al., 2019: 25). The only thing that may be considered surprising is that the other ethnic groups do not display significant results.

Figure 1 displays the magnitude of the effects for all the demographic variables. For visual representation, the effect sizes have been scaled, and the lines present the 95% confidence intervals. An odds ratio of 1 means no effect change. As can be seen, the main positive effects come from the class and education variables, while the control variables display little (negative) effect, with the only exception being the white ethnic group, which has a significant and sizeable negative effect on a person's propensity to support Labour.

With regards to all the models presented here, the model fits for these demographic-only models are quite low. This indicates that looking at individual-level characteristics only tells part of the story on why people do and do not support Labour. This means that more is needed to explain the variation of support for Labour. The contextual models below will help to elucidate more.

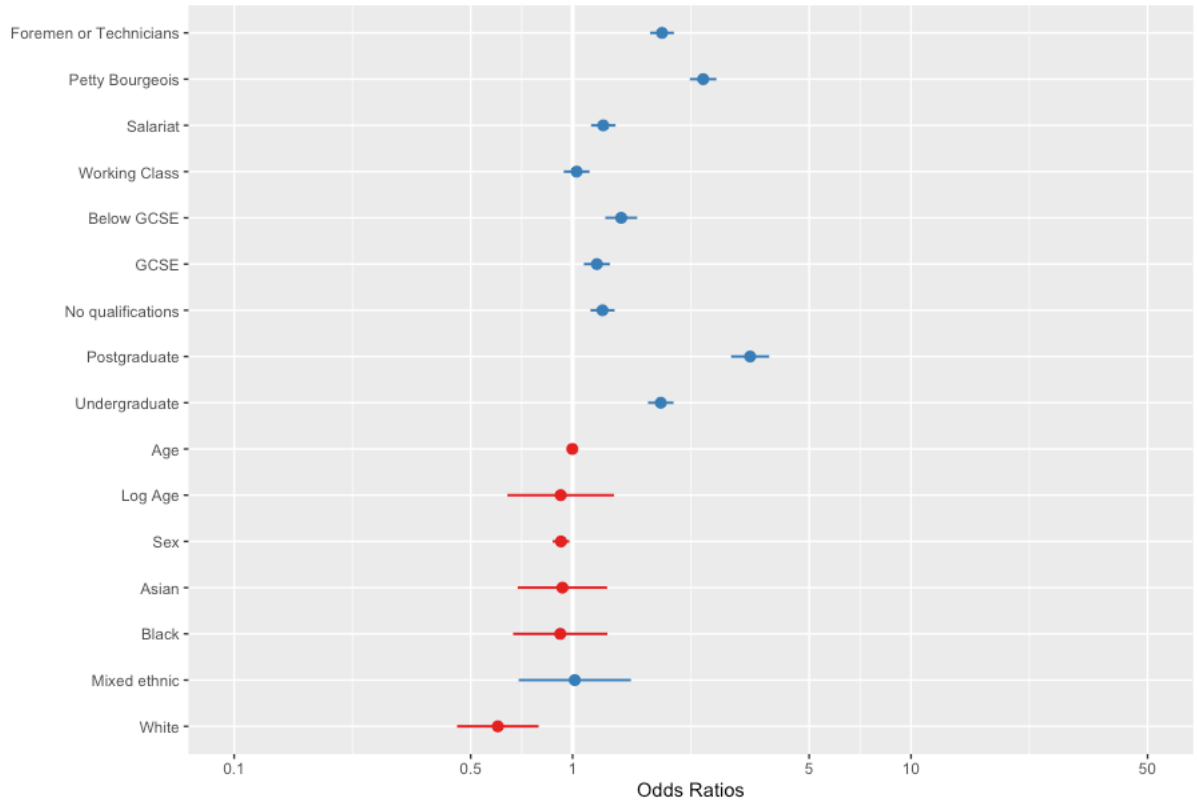
Bringing in the context: individual-level only

Just like Arzheimer and Carter (2006: 429-33), I present my multilevel model by first showing the results of the individual-level characteristics (level 1) in the survey clusters, before introducing the contextual variables (level 2). Table 3 displays both models.

As can be seen in Table 3[13], placing the individual-level variables in the survey clusters leads to changes in the results. While all social class variables that were significant do remain significant, the impact for most decreases. While only the Petty Bourgeois category actually displays a lower significance in terms of p-value, the "Foremen or Technicians" and "Salariat" categories show a smaller effect size. In fact, where the Salariat class used to have a slight positive effect, it now becomes slightly negative.

The most pronounced change, however, occurs among the working class. The odds ratios now indicate that working class has a significant effect; the effect size now is highly significant positive. This means that, when factoring in potential differences between different surveys, a person's working-class identity has a significantly positive effect on their propensity to support the Labour party. In fact, with the different surveys accounted for rather than just pooled into one group, a traditional class-support base for the Labour party starts to emerge.

Figure 1: Base demographic model



Note: in this figure, red refers to a negative effect, and blue to a positive effect.

[13] This model is slightly simplified for visual representation. The full model is included in Appendix D.

Table 3: individual-level model and full model

Dependent variable: Labour support

	Individual-level only model	**Full model**
Predictors	*Odds Ratios*	*Odds Ratios*
(Intercept)	0.00 ***	0.00
Foremen or Technicians	1.39 ***	1.16
Petty Bourgeois	1.16 **	0.64 **
Salariat (Higher & Lower)	0.81 ***	0.80 **
Working class	1.60 ***	1.41 ***
Below GCSE	1.19 ***	1.03
GCSE	1.10 *	1.09
No qualifications	1.53 ***	1.49 ***
Postgraduate	1.30 ***	1.09
Undergraduate	1.16 ***	0.96
Age	0.99 ***	0.81 ***
Log Age	1.33	1.32
Male	0.98	0.96
Asian or Asian British	2.64 ***	1.71
Black, African, Caribbean or Black British	3.33 ***	2.34 **
Mixed or Multiple ethnic groups	1.87 ***	0.93
White	1.16	0.72
Labour government		3.47 **
Party system polarization		1.97
Complexity		0.56
Welfare state complexity		2.75 ***
Random Effects		
σ^2	3.29	3.29
N	29 survey	29 survey

Observations	77605	77605
Pseudo R^2	0.00	0.07

** p<0.1 ** p<0.05 *** p<0.01*
Note: reference categories are Clerical (junior non-manual) for social class, A-levels for education, 18 years for age, female for gender, and Other ethnic (including Arab) for ethnic group.

This is most clearly pronounced by the difference between the "Salariat" and the working class. While the salariat displays a significant negative effect, the working class displays a significant positive effect. Drawing back on traditional class divisions and views, we can see that the class self-image of the salariat was overwhelmingly "middle class", while the lowest occupational grades (the skilled and unskilled manual workers) were the ones who most identified with the working class (Butler and Stokes, 1977: 72). What this most clearly tells us, is that the traditional class-based divide in support for the Labour party still has explanatory power. As can be seen by just looking at the individual-level model, the divisions between the higher and lower classes in terms of their support for the Labour party still exist.

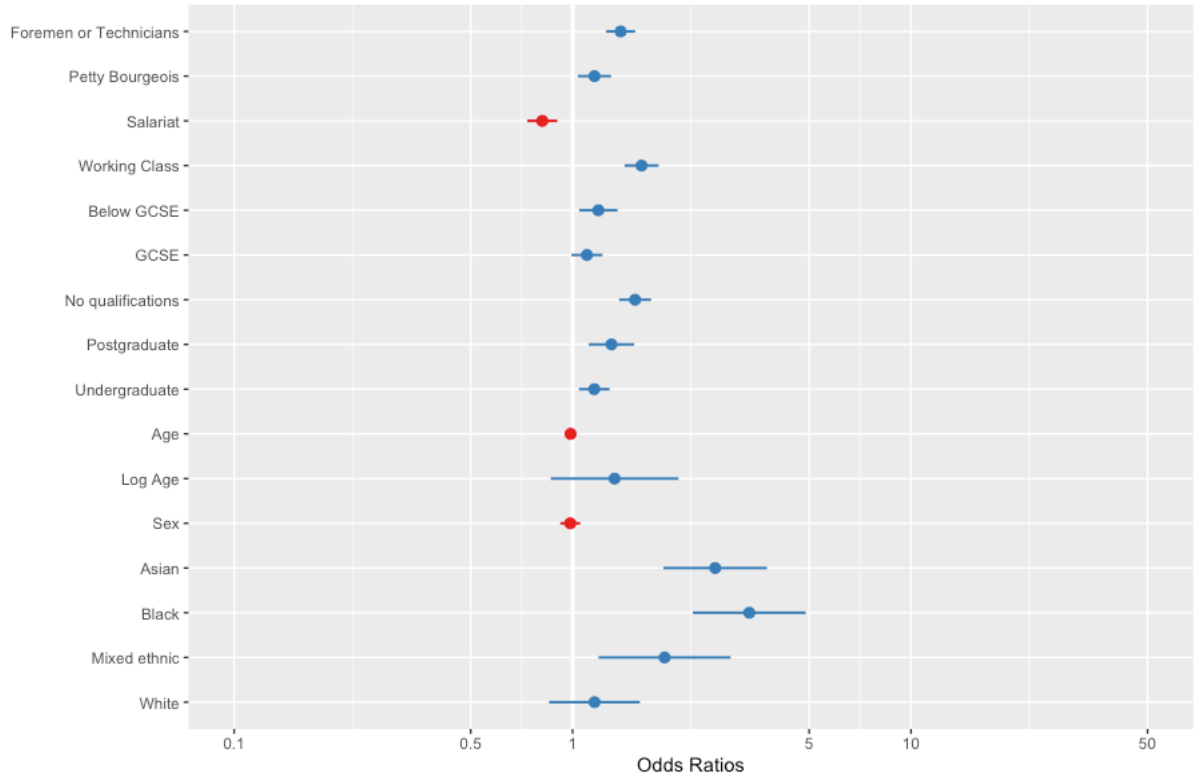
First, though, let us go through the rest of the individual-level model to see the results. As before (in Table 2), all the education categories have a (highly) significant positive effect on a person's propensity to support Labour. Except the GCSE-educated individuals, that is. Where, before, the GCSE-educated group showed a slightly positive effect, this effect is now smaller, and less significant. It seems that the people who have graduated with a GCSE degree and then stopped their education journey are the least likely to support Labour.

With regards to the other education groups, the effects remain positive, but less overwhelmingly so, compared to the models in Table 2. In fact, while university-educated individuals are still positively inclined to support Labour, they are no longer the group with the most positive effect. This spot is now reserved for another part of the traditional Labour constituency: the individuals with no formal qualifications. Here again, like with the class-support, the model seems to show that the traditional Labour base is the most promising group in terms of support for Labour, rather than the newer, higher-educated, middle class groups that have been at the fore of Labour strategic campaigning since the mid-1990s (Dalton, 2021).

Figure 2 shows the odds ratios visually, in scaled form for easier visual interpretation. This confirms the abovementioned image that a class divide in support for the Labour party can be distinguished, as well as smaller education-level effects.

Going over the control variables, we can see that age again has a small negative effect, with an odds ratio of 0.99. That said, as age invariably changes over time, the build-up of this small negative effect can, over decades, lead to a stronger negative effect. What this result tells us, therefore, is that people become very slightly less inclined to support Labour as they grow older.

Figure 2: Individual-level model with survey clusters



Note: in this figure, red refers to a negative effect, and blue to a positive effect.

This is in line with what has been found in the literature, with British voters becoming "substantially" more likely to vote Conservative when they grow older (Tilley and Evans, 2014). Meanwhile, the log transformation of age again has no significant effect.

With regards to gender, the effect is almost as small as with age, and no longer significant. This means that there is no discernably significant difference in the propensity to support Labour between female and male respondents.

Ethnicity, on the other hand, provides a different story altogether. No longer do all but one category show no significant effect; quite the contrary, all but one show a significant effect. Moreover, all categories display a positive effect, the largest being among Black and Asian individuals. The only non-significant category is White, which is noticeably different from the demographics-only models in Table 2. It seems that, when taking different survey clusters into account, the effect of identifying as white no longer has a significant effect on a person's likelihood of supporting Labour.

As we can see therefore, just the introduction of the survey clusters already seems to have a conditioning effect which affects the individual-level variables. When the contextual variables are introduced, their influence becomes even more pronounced, confirming that the context has an impact on the individual.

Contextual models: introducing the contextual variables

As we can see in Table 3, model named "Full model" includes the contextual variables of interest to this study. Furthermore, Figure 3 shows the effects in the full contextual model, compared to the level-1-only model. With these results, hypotheses 3 through 6 can be assessed.

The first thing that becomes immediately noticeable when just looking at the contextual variables, is that their effects vary wildly. To look at the contextual control variable, government status, we can see that there is a significantly strong positive effect on a person's propensity to support Labour when this party is serving in the government. In other words, when Labour is in charge, people will more likely to support the party. This is suprising, as the expectation was that people tend to be turned away from the party in charge, for instance during second-order elections (Heath et al., 1999: 394). That said, that people are supportive of the party in charge can also be indicative of a positive incumbency effect (Peskowitz, 2017).

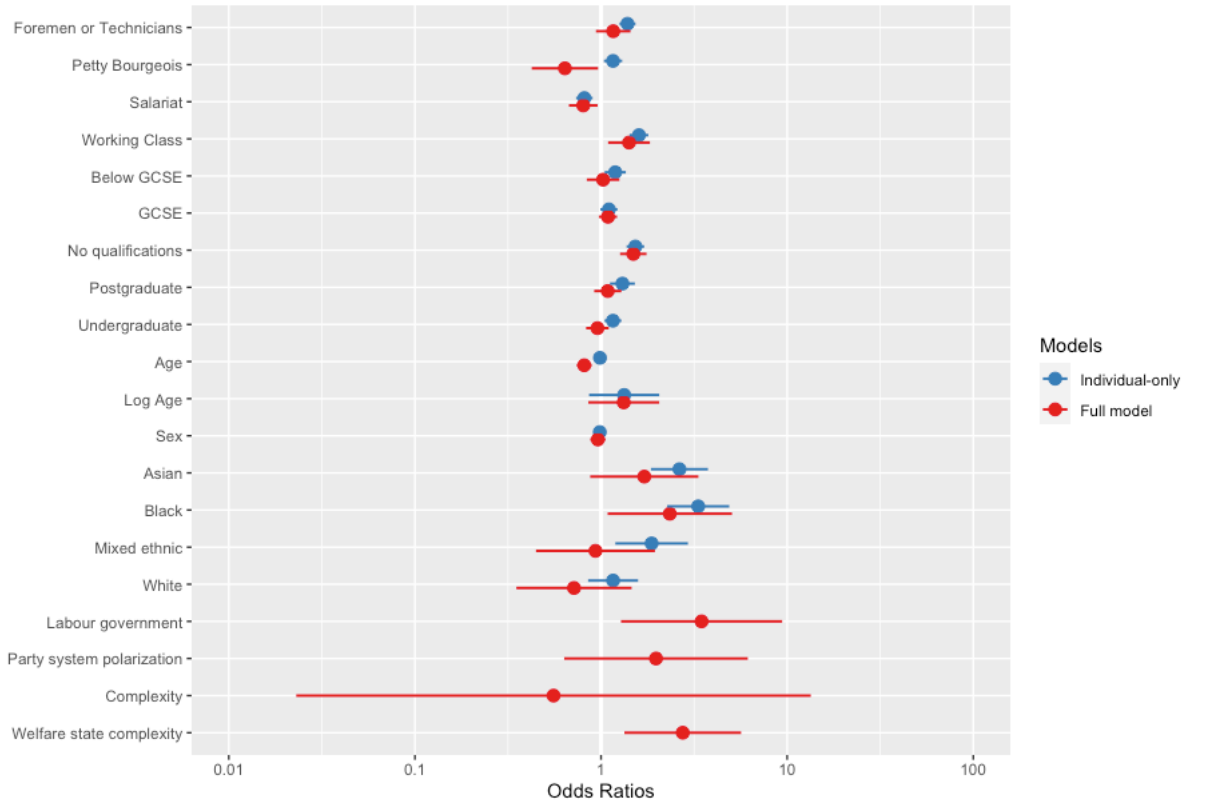
With regards to the other supply-side explanation derived from the literature, the party system polarization (measured using Dalton's polarization index), here again a clear positive effect can be seen. This means that when there is *more* polarization in the British party system, people are *more* inclined to support the Labour party. However, the results are not significant. This can indicate that polarization itself does not (dis)incentivize people to support Labour; rather, it

provides the context for individual-level factors. Therefore, the effect of individual-level characteristics varies depending on different contexts.

A large negative effect appears for language complexity. What this means is that when the language spoken in Parliament by all speakers becomes *more* complex, people will be *less* likely to support the Labour party. That said, the effect *in itself* is not significant. This means that we cannot say that variations in language complexity used across speakers has a significant effect on voters. But that is not what this investigation is looking for. Similar to polarization, there is no assumption of any clear increase or decrease in support for Labour caused by variations in language complexity *eo ipso*. Rather, with both these factors (polarization and language complexity), the expectations are that they will *influence* the effects of the individual-level characteristics that shape a person's propensity to support Labour.

When we take these factors into account, a more nuanced picture comes to light. Starting with class, there is a clear difference between the previous, "context-less" model and the one including the contextual variables. No longer are all classes significantly impacting Labour support; rather, only three class categories remain significant. In line with the individual-level-only model in Table 3, the traditional Labour class-support base re-emerges. With the

Figure 3: Individual-level model and full model



"Salariat", petty bourgeois and working-class now significant, two traditional class-based groups have emerged. Figure 3 presents the odds ratios of the full model placed on top of the individual-level-only model, in order to display the differences in effect size that appear when the contextual variables are included.

First, the salariat. These people are the ones in leadership (managerial) positions. They give directions to others, review performance, manage team or company budgets (Goldthorpe, 2016: 90). They are distinctly not *workers* in the sense that they are not simply told what to do– rather, they are the ones giving orders.

Second, the petty bourgeois. Like the salariat, they are not the workers, they are the ones setting out the policies. However, they do not do so in managerial positions, but in their own shops and on their own farms. They are distinctly no *workers* – they do not sell their labor power to an employer (Steinmetz and Wright, 1989: 979). Rather they are "self-employed" and may or may not hire workers. As such, the petty bourgeois here can be grouped together with the salariat here.

Third, the working-class group. These people are, economically speaking, the polar opposite of the salariat (and, to a lesser extent, the petty bourgeois). They do not give directions to others, they *are directed by others*. They sell their labor power to an employer for a wage (Steinmetz and Wright, 1989: 979). The economic experiences of these two groups, are markedly different.

Traditionally, these two groups were not only economic polar opposites, but politically, too. In 1970, 80% of the salariats in Britain considered themselves middle class, while 80% of the workers considered themselves working class (Butler and Stokes, 1977: 72). This *class* self-image was transformed into a *partisan* self-image, with overwhelming numbers of middle class-citizens identifying as Conservative, and a similarly large part of working class-citizens identifying as Labour supporters (Butler and Stokes, 1977: 77). As discussed in the introduction and theory chapters, this clear link has largely disappeared over time in Britain.

But what the results here show is that these two groups are still available, and relevant, *given the right context*. As can be seen in Table 3, the salariat and petty bourgeois are significantly *less* likely to support Labour *in the context* of party system polarization and complex language. At the same time, the working class is significantly *more* likely to support Labour under the

same contextual conditions. As such, the results confirm the expectations drafted earlier on in this thesis. Namely, that *context activates individual-level characteristics that define support*. In other words, context matters, because it activates the individual characteristics that drive a person's support for the Labour party.

This means that both hypothesis 3 and hypothesis 4 are supported. After all, what can be seen is that, when the contexts of party system polarization and language complexity are introduced, the most pronounced class-based difference in party support (between the managers and the workers, the traditional middle and working classes) become key again. The key takeaway from this seems to be that once we control for ideological polarization and language complexity, individual-level characteristics like social class become more important in defining party support. As such, the context provided by polarization and language matters.

In light of this, we can also assess hypothesis 5 on the influence of context on the education-based support for Labour. Where, before, most education categories had significant influence on a person's propensity to support Labour, most of these categories are now no longer significant. Rather, the only significant category left is the group without formal qualifications. This group is not only the only group with a sizable effect, but also the only group which has a significant effect. This means that when the language use in Parliament becomes *more* complex, lower-educated individuals become *more* likely to support Labour.

This is in line with the class finding. After all, just like with the class support, the traditional core constituency of the Labour party, the lower-educated people, are more likely to support their traditional party under these conditions. When the language used in politics becomes more difficult, lower-educated people go back to what they know and support Labour. As such, this finding is in line with the expectations.
When assessing hypothesis 5, however, a bit of a conundrum emerges. Yes, lower educated people are moving towards Labour under these conditions, but *higher* educated people are not equally turned away from Labour in the same context. Therefore, it cannot conclusively be argued that more complex language in politics leads to more education-based support for Labour. As such, hypothesis 5 cannot be confirmed.

With regards to debates on the welfare state, the expectations formulated earlier were that these have a positive effect on a person's propensity to support Labour. This means that, in the context

of debates on these topics, more complex language in Parliament actually helps, rather than hinders Labour. Where it was shown earlier that the *overall* language complexity *eo ipso* has a negative, but not significant effect on Labour support, the expectation was that the results will be different when the language becomes more complex in *specific* contexts – certain debate topics, like the welfare state.

As becomes clear from the results, this expectation is supported by the results. Language complexity in the debates on the topic of the welfare state has a highly significant and strong positive effect on a person's likelihood of supporting Labour. This makes intuitive sense. After all, topics related to the welfare state go to the very core of social democratic party policy – these parties care a great deal about using state mechanisms to enhance the welfare of citizens (Przeworski, 1980: 52). As such, assuming that these parties, and therefore Labour as well, are "issue owners" on topics related to the welfare state (Budge, 2015), it makes intuitive sense that they are considered "experts" in debates on this topic. When debates on this topic will then involve more complex language, the natural inclination will be for people to turn to, in Britain, the Labour party. This comes back to the basic idea of representation presented earlier: who can accurately *represent* you. On the topic of the welfare state, the results show clearly that people will be much more likely to answer "Labour". Therefore, given that the propensity to support Labour is strongly positively affected by more complex language in welfare state debates, hypothesis 6 can be confirmed.

Conclusion and outlook

In this chapter, a big step was taken towards more understanding about the influence of language complexity on Labour support. In the process, it became clear that context matters.

The results showed that, without context, it seemed that the higher-class, higher-educated constituents were more likely to support Labour. It seemed that the working class was no longer a relevant factor in determining Labour support. Furthermore, those identifying as white were much less likely to support Labour. It seemed, therefore, that the traditional working-class base of the Labour party was no longer a relevant support factor. With the working class not being significant as a determinant for Labour support, and lower education levels displaying smaller levels of support, it seemed that the traditional Labour core constituency had indeed been

replaced by the new electorate, the middle-class, higher-educated individuals so clearly appealed to by New Labour (Dalton, 2021: 1).

Enter contextual factors. When these were added, the whole picture changed. It became clear that when the contextual factors of polarization and language complexity were added, traditional support bases for Labour re-emerged. The middle-class voters were turned off from supporting Labour, while the working class was significantly more likely to support their traditional party, as was the lower-educated group. This factor was even more clear in debates on a core "Labour policy", the welfare state, which showed in a strong increase in support for Labour.

What does this mean for the language recognition thesis? The preliminary conclusion would be that the relationship is supported. People are looking for representation by someone who actually *speaks* to them. If the language used by politicians becomes more complex and therefore cannot accurately signal what politician to support, it then becomes natural to return to the party that *feels* closest. In case of working-class and lower-educated citizens, this turns out to, still, be Labour. More so, when the topic under discussion concerns issues related to the welfare state, something which Labour is associated with, complexity is no longer an issue – what Labour will say will be reliable, whether it be difficult or not.

While these initial conclusions seem to confirm the language recognition thesis, some questions remain. For instance, when language complexity is the only factor that matters, the only thing that changes, does it then still activate class divisions? And if *more* difficult language brings back traditional class divisions, what happens if a party starts using *less* difficult (or even outright simple) language? Can, for instance, a Conservative politician snatch away the working class vote, simply by adopting simpler language? To investigate this, extra investigation is needed. For that purpose, I conducted a survey experiment. In the next chapter, I will outline my experimental design and the choices that I made in assessing the results. The chapter after that presents the results, in order to draw more conclusions towards a better understanding of the mechanism proposed by the language recognition thesis.

Methodology 2: explaining the experiment design

Having discussed the results of the regression, this chapter will explain the second step of the research strategy: the experiment I use to investigate the proposed mechanism. As this thesis is interested in how people perceive political language use, using an experiment to subject people to political language of varying complexity is a natural method of exploring how people respond to variations in language complexity. As such, the experiment provides a welcome addition to the regression model that was discussed in the previous chapter. I use the experiment mainly to investigate hypothesis 7, which posits that parties can "lure" away the traditional class base of the Labour party, by virtue of using simpler language, *irrespective* of what party they represent.

Through the experiment, I am able to draw stronger inferences based on the fact that experiments help increase internal validity of the argument. Furthermore, as the participants will read texts of varying difficulty (see description below), this allows me to present purely the texts of varying difficulty to the participants, without them being influenced by whoever is uttering the texts, at what tone, representing what party, and via what medium they hear about the text (as is the case with individuals normally being exposed to parliamentary speeches). Moreover, by not informing the respondents of the speaker's identity, I rule out any influence of gender, sexual orientation, age or ethnicity on the participants' response to the texts.

Furthermore, the measure used to determine language complexity, the Flesch-Kincaid Grade Level score, is made for readability (Kincaid et al., 1975). Therefore, having the participants read the texts of varying difficulty most closely aligns with the purpose of the measure used. As the preparation of the dataset from which speeches are drawn has been described already in chapter 3, I shall not repeat description of this process here, and instead will focus purely on the experimental design. The experiment is described below.

Outline and sample selection

I conducted an online survey experiment, using randomly selected participants from the subject pool of the Nuffield Centre for Experimental Social Sciences (CESS). A sample of N=328 is used for the analysis. This is close to the desired sample size of N=333 that was written down prior to data collection in my Pre-analysis Plan (Appendix E).

The reasoning behind this sample size is both practical and calculated. Practical, because of the financial constraints to the study (the available funding combined with the desire to reasonably compensate all participants for their participation). Calculated, through a sample size calculation based on the desired effect size, which was taken from a piece of work on a similar topic (Robison et al., 2021). With the desired effect size, I used a two-sampled t-test, with statistical power of 0.80 and a significance level of 5%, to determine the minimum required sample size per group at N=16. Given that my experimental design includes two treatment factors (see below), a minimum total sample size of 16 x 2 x2 = 64 would be required. Increasing the sample size to N=328 (82 per group) allows me to do sub-group analyses based on the responses to the social characteristics questions that are asked after the experiment.

While subjects are randomly drawn, there is one selection criterium. The subjects need to be British citizens, to make sure they are eligible to vote in Great Britain. This is because the study investigates the influence of language use in British politics on class voting behavior in Britain.

Each participant in the study reads four short political speeches (maximum 200 words each), taken from the parliamentary speeches dataset, based on the *Parlspeech V2* dataset (Rauh and Schwalbach, 2020) I use as the basis for my thesis (see selection criteria in chapter 3). None of the speeches are about proposals that are currently being voted on, and participants are not being told what political party the speaker represents. This is to prevent outside influence on the lab conditions. The number of four speeches is to allow every participant to receive both the treatment and control of both treatments in the experiment (see below), and to allow the participants to complete the study in 15 minutes.

All speeches were taken from debates on unemployment benefits, which is a topic closely related to social class. The texts were selected after reviewing the results of the topic modelling exercise (see chapter 3). The reason I selected texts from only one debate topic is that I want to prevent topic-based effects leading to varying results in my data. Selecting more topics could have been a strategy worth pursuing, but the financial constraints to the study did not allow for this. Therefore, investigating the mechanism across different debate topics is left for future experimental research. In this study, however, by limiting the experiment to just one debate topic, I can safeguard the internal validity of my findings.

After the experiment, the participants are asked to answer six personal characteristics questions, relating to their social class, the social class they grew up in, their education level, their parents' education level, and their political affiliation.

Treatment

The assignment of treatments to subjects is randomized. However, I use a within-subjects design, which means that all subjects receive all treatments and controls. The order of the treatments, and which texts are treated is randomized, however. This allows me to control for ordering and treatment selection effects.

The experiment contains two treatments. The first treatment is *simple text*, versus *difficult text* as the control. *Simple text* here means a lower Flesch-Kincaid Grade Level score, as described earlier in this chapter, and *difficult text* a higher score (Kincaid et al., 1975). For the experiment, I have taken speeches of the difficulty level "difficult" (corresponding to the reading proficiency of a Bachelor-level university student) and, for the simple text, "standard" (corresponding to the reading proficiency of a 13 or 14-year old school student).

The second treatment allows me to prevent a confounding variable in the form of *altered text* versus *original text*. For this treatment, I simplified difficult texts to the Flesch-Kincaid score of "simple text" (as defined above), and I made the simple texts more difficult. The reason for this is that I wanted to see what the effect would be if the same text would be difficult versus if it would be simple. By taking the treatment of *original* versus *altered*, I can determine whether there is an effect of altered versus original text, and, if so, control for it in my analysis.

Controls

Of the four texts presented to each participant, two are simple, and two are difficult. Furthermore, two are original, and two are altered. Therefore, each participant will read one of each text of the following characteristics: *simple-original*, *simple-altered*, *difficult-original*, *difficult-altered*. The participants are informed in their instructions that the texts they read contain original and altered texts. They are not informed that the texts are of varying difficulty, as this would defeat the purpose of the experiment, and could lead to participants explicitly looking for which text is difficult, and which is not.

To control for the effect of specific texts, I have created a dataset with multiple difficult and simple texts, in original and altered form. By randomizing which texts are presented to the participants, and in what form, I remove the effect of specific texts.

A second control I include is a control for potential ordering effects. This is done by randomizing when participants are presented with what type of text. So, while one participant may view a *difficult-altered* text as the first text, a different participant may view a *simple-altered* text first. By randomizing the order in which texts and treatments are presented, I can account for any ordering effects that may have occurred otherwise. An example experiment is included in Appendix F.

Analysis of test results using proportions test

After reading each text, the participants receive three questions. These variables are *political distance* ("Do you feel this person understands people like you?"), *political support* ("Would you vote for this person?"), and *speaker-party affiliation* ("What political party do you think this person is affiliated with?"). The main variable of interest is the *political support* variable, as this allows me to effectively assess hypothesis 7. After all, hypothesis 7 is supported, we should expect to see significantly more respondents answering "Yes" to this question than "No" when a Conservative politician uses less complex language. The *political distance* variable helps assess hypothesis 7 as well, as people may not vote for a politician, but still feel that the politician understands people like them.

The third variable is a control variable, together with the *party support* variable taken from the post-election survey questions. Combining these results allows me to control for effects based on a participant's personal political preference (not) aligning with the speaker's party affiliation. In other words, this is a control measure to account for the possibility that, for instance, Labour party voters do not want to vote for politicians they think are Conservative.

I measure the treatment effects by using a proportions test (2-sample test for equality of proportions). I take a baseline value for all the four hypotheses investigated using the experiment, by looking at what all the participants answered across the questions for the *political distance*, *political support*, social class, education level and party identification

variables. I then compare the proportions according to the treatments received, to determine the treatment effects. I then look at the differences between the effects when the speaker is a Labour politician versus a Conservative politician, in order to assess hypothesis 7.

Conclusion and outlook

In this chapter, the experimental design was discussed. It was noted that two treatments are introduced to the participants: simple versus difficult text, and original versus altered texts. The presentation of which texts are simple and which difficult, as well as which are original and which are altered, is randomized, as is the order in which texts are presented. As has been described, the results are analyzed using a proportions test.

The next chapter will present the experimental results, and give the opportunity to test hypothesis 7, before moving on to the conclusion chapter.

Results 2: experimental results show respondents more likely to support politicians using simpler language

Having discussed the design of the survey experiment, I now turn to the analysis of the results. In this chapter, I start by describing the original data that was gathered through the experiment. After this, and in line with the Pre-analysis Plan (PaP, see Appendix E), I start by taking baseline values of the differences between simple and complex speeches. Doing so will help give a clearer idea of whether or not using less complex language in parliamentary speeches can have a positive effect on the respondent's feelings of distance towards the speaker, as well as their willingness to vote for the speaker. Here, by means of control, the difference between original and altered texts will be investigated as well, in order to make sure no significant differences exist between the texts originally spoken in Parliament, and those altered for the experimental treatment.

After this, the results will be viewed in light of the respondent's self-reported social class status, education level, and which party they support. Then hypothesis 7 will be assessed, by looking at class-based results for simple and difficult speeches by both Labour and Conservative politicians.

The results will be interpreted using a two-proportions z-test, which gives a clear indication of the significance of the differences observed. Contrary to what was, erroneously, written in the Pre-analysis Plan (Appendix E: 3), I shall not use a significance level of $p=0.5$ as the threshold for significance. As this would not give enough confidence in the results, I shall take the stricter levels of $p<0.1$, $p<0.05$ and $p<0.01$ as the significance benchmark to test the results below, as these are the same significance levels used in the regression chapter and in other studies (such as Jansen et al., 2013: 58).

Description of the data and comparison with the general population

In total, 328 individuals participated in the study. Contrary to the dataset used in the regression chapter, the experiment dataset does not contain continuous variables (only categorical ones), and therefore presenting a standard descriptive statistics table (including figures like a mean, median and standard deviation) is not possible. That said, some statistics about the respondents can be presented here, and this is done in Table 4. This table shows what type of people the group of respondents was made up.

Table 4: Descriptive information of respondents

Social class		**Education level**		**Party identification**	
Answer given	*Frequency*	*Answer given*	*Frequency*	*Answer given*	*Frequency*
Middle class	108 (34.50%)	Undergraduate degree	125 (38.46%)	Labour	99 (31.13%)
No class	70 (22.36%)	Postgraduate degree	97 (29.85%)	A mixture/somewhere between various parties	46 (14.47%)
Working class	56 (17.89%)	Some university	60 (18.46%)	Liberal Democrat	34 (10.69%)
Lower middle class	56 (17.89%)	Secondary school	41 (12.62%)	Conservative	33 (10.38%)
Upper middle class	12 (3.83%)	Prefer not to say	2 (0.62%)	Green Party	30 (9.43%)
Lower class	6 (1.92%)			No/none	18 (5.66%)
Prefer not to say	5 (1.60%)			Central/somewhere in the middle	17 (5.35%)
				Any party I agree with at the time	17 (5.35%)
				Prefer not to say	10 (3.14%)
				Independent	7 (2.20%)
				Scottish National	4 (1.26%)
				United Kingdom Independence Party (UKIP)	1 (0.31%)
				Brexit Party	1 (0.31%)
				Other	1 (0.31%)
N	313		325		318

First, looking at the social class that the respondents indicated they identified with. 313 respondents answered this question (15 did not). The largest group (34.50%) identified as middle class, while a little over a fifth of the respondents (22.36%) did not identify by class. 17.89% of the respondents identified as working class, and an equally large group as lower middle class. Upper middle class (3.83%) and lower class (1.92%) were the smallest class groups. 5 respondents (1.60%) refused to answer.

Where this respondent's pool seems to identify more with the middle class and with no class than with the working class, this is different from what other studies have found. The British Social Attitudes survey (waves between 2003 and 2015), for instance, found that about 25% of the respondents self-identified as working class, and about 20% as middle class (Evans and Mellon, 2016: 4). Furthermore, the number of people who identified with no class whatsoever was much larger at around 50% (Evans and Mellon, 2016: 4).
What we can say about the respondents in this experiment, then, is that they identify more as middle class, and less as working class, than the general UK population.

As can be seen in the education columns, the respondents were, relative to the general UK population, highly educated. Of the 325 participants who answered this question, more than a third (38.46%) has at least an undergraduate degree, and a little under a third (29.85%) has a postgraduate degree. Together, this means that a total of over two-thirds of the respondents (68.31%) has at least one university degree. This is much higher than the UK national average, which stands at about a quarter of the population having an undergraduate degree (25.5%) and only 1.5% having a postgraduate degree (OECD, 2021). As such, the respondents in this dataset are, relative to the UK population, highly educated. The people who attended university but did not receive a degree (18.46%) and those who had as highest qualification a secondary school degree (12.62%) complete the respondent pool. Two respondents refused to answer, while three respondents did not answer.

Finally, party identification. A little under a third of the respondents (31.13%) answered "Labour" as the party they identified with. As such, this was more than any other party. The other parties that were mentioned were Liberal Democrat (10.69%), Conservative (10.38%), Green Party (9.43%), Scottish National (1.26%), UKIP (0.31%) and the Brexit Party (0.31%). The number of respondents who did not identify with any specific party or ideological direction (such as "Central/somewhere in the middle") is noteworthy: in total, about a quarter (27.67%) of the respondents classified themselves as "non partisans" in this way.
Comparing this to the UK general public, it becomes clear that while the Labour support base is comparable to that among the general public, the "non partisan" group is slightly bigger, and the Conservative group is smaller than the support share among the general public (Lee and Young, 2013: 65). As such, the Labour support is relatively more prevalent in this respondent group than in the UK population.

Baseline differences conditioned by the treatment and control treatment

In line with the analysis proposed in the Pre-analysis Plan (Appendix E: 3), I now first analyze the baseline differences on the treatment of simple versus difficult speeches, and the control treatment of original versus altered speeches. Tables 5 and 7 show the proportions of respondents answering "Yes" versus those answering "No" on the questions "Would you vote for this person?" ("political support") and "Do you feel this person understands people like you?" ("political distance"), presented for the treatments of simple versus difficult speeches (Table 5) and original versus altered speeches (Table 7).

Table 5: Porportions table of simple versus difficult texts

Variable	*Yes*	*No*
Political support: simple	0.609	0.373
Political support: difficult	0.534	0.448
Political distance: simple	0.649	0.333
Political distance: difficult	0.577	0.405

As can be seen by simply looking at the tables, the proportion of "Yes" answers on both the political support and the political distance variables is higher among the simple texts than among the difficult texts. This gives tentative reason to be optimistic. However, just looking at the tables of course cannot be sufficient; a test needs to give confirmation or reject the idea that the effect of language complexity matters here.

Table 6: Proportions test results simple versus original treatment

Variable	*Chi-squared*
Political support	2.9276*
Political distance	2.7269*

** p<0.1 ** p<0.05 *** p<0.01*

Therefore, using the aforementioned two-proportions z-test, I tested whether the difference between simple and difficult texts in terms of how many people answered "Yes" to the political support and political distance questions was significant. The results are shown in Table 6. The proportions test for the political support variable gave a chi-squared value 2.9276 and p-value

of 0.08708. Seeing that 0.08708<0.1, we can conclude that the effect here is significant at the p<0.1 level.
Likewise, with the political distance, the proportions test delivered a chi-squared value of 2.7269 and p-value of 0.09867, which, again means that the effect is significant at the p<0.1 level.

One potential confounder that was identified before the experiment was conducted, is the potential influence of using original speeches (i.e., the ones actually spoken by MPs in Parliament) versus using altered versions of those speeches (which allow for investigation of varying language complexity). Therefore, this was investigated as well. As Table 7 shows,

Table 7: Porportions table of original versus altered texts

Variable	*Yes*	*No*
Political support: original	0.596	0.404
Political support: altered	0.549	0.436
Political distance: original	0.622	0.378
Political distance: altered	0.600	0.385

the proportions of respondents answering "Yes" on the political support and political distance are relatively similar for the original and altered groups. This gives a first indication that respondents did not have a markedly different experience reading either original or altered speeches.

Using the same two-proportions z-test as above on first the political support question results in a chi-squared value of 1.0699 and p-value of 0.301. As 0.301 is greater than 0.1, the conclusion here must be that the null hypothesis cannot be rejected. In other words, there is no significant difference between original and altered speeches (at least not at the p<0.1 level) for the political support question. The results are presented in Table 8.

Table 8: Proportions test results simple versus original treatment

Variable	*Chi-squared*
Political support	1.0699
Political distance	0.19123

** p<0.1 ** p<0.05 *** p<0.01*

On the political distance question, a similar picture emerges. With a chi-squared value of 0.19123 and p-value of 0.6619, it again becomes clear that there is no significant difference (at the $p<0.1$ level) between the original and altered speeches' effect on a respondent's likelihood of answering "Yes" in response to the political distance question. This gives clarity on the effect of altering the complexity of the speeches: there is no significant difference. As such, we can be confident that the effects of language complexity are not confounded by the difference of original versus altered speeches.

To quickly recap the findings here: there is a significant effect of language complexity on a person's likelihood to vote for a politician (political support) and a person's likelihood to feel understood by a politician (political distance). As such, the initial intuition of the language recognition thesis – that language complexity can impact a person's feeling of recognition with a politician – is supported by these findings.

One caveat must be added, however: while the results are clear, it must be remembered that the experiment only used parliamentary speeches from one debate topic (welfare benefits). Therefore, it must be remembered that all conclusions in this chapter on the effects of language complexity should be drawn while bearing in mind that the certainty of the findings does not necessarily travel beyond speeches on this topic. In other words, a finding of a significant effect here does not necessarily mean that the effect will be significant in debates on other topics as well. Future work should explore whether the same effects can be found among debates on other topics.

Nonetheless, the initial results are encouraging. Below, the results conditioned on respondent's personal characteristics are presented, starting with self-identified social class, then highest education level achieved, and ending on respondent's partisan identification.

Differences per social class categories

When looking at the social class categories, we can see that six categories (no class, middle class, working class, lower middle class, upper middle class, and lower class) are among the chosen options by the respondents in this experiment. The five respondents who refused to answer cannot be included in the analysis here, as it cannot be determined what their class identity is. Furthermore, as only six respondents selected "lower class" as their class identity,

conducting any meaningful proportions test on this class category is not possible. Therefore, the analysis here will focus on the five other class categories. The results of the proportions tests are presented in Table 9. What these tests tell us, is whether there is a significant difference among the proportions of those, *per social class category*, that answered "Yes" to a question about a simple speech compared to the "Yes"-answers in response to difficult speeches. If there is, we can determine that there is a difference of language complexity effect between different social classes.

Starting with those who selected "no class", the proportions tests yielded values of 2.4874 (chi-squared) and 0.1147623 (p-value) for political support, and 0.23123 (chi-squared) and 0.6306164 (p-value) for political distance. We can conclude that there is no significant difference between simple and difficult speeches among the "no class" group.

Table 9: Proportions test results per social class

Category	*Political support: chi-squared*	*Political distance: chi-squared*
No class	2.4874	0.23123
Middle class	0.087662	0.75006
Working class	0.48093	1.1473
Lower middle class	0.018574	0.28592
Upper middle class	0.33862	0

** p<0.1 ** p<0.05 *** p<0.01*

With regards to the middle class, the proportions tests resulted in values of 0.087662 (chi-squared) and 0.7671703 (p-value) for political support, and 0.75006 (chi-squared) and 0.3864559 (p-value) for political distance. Again, we can see that no significant effect of language complexity can be distinguished for this group.

For the working class, the proportions tests resulted in values of 0.48093 (chi-squared) and 0.4879994 (p-value) for political support, and 1.1473 (chi-squared) and 0.2841069 (p-value) for political distance. Therefore, varying language complexity cannot be considered to have a significant effect on the working class group.

In the lower middle class group, values of 0.018574 (chi-squared) and 0.8915954 (p-value) for political support, and 0.28592 (chi-squared) and 0.5928474 (p-value) for political distance again do not display significant effects of language complexity. The same can be said of the upper middle class group, on both political support and political distance. Especially the political distance variable among the upper middle class group shows that there is no significant effect of language complexity – the proportions of those answering "Yes" are the same for simple as for complex language.

What this shows is that, when the effect of language complexity on political support and political distance is conditioned on a respondent's social class identification, no significant effect arises. In other words, no class-based effect can be found on the influence of language complexity on political support and political distance.

Differences per education levels

With regards to education, four categories can be used in the analysis (undergraduate degree, postgraduate degree, some university, and secondary school). The respondents who refused to answer or did an answer are omitted from the analysis. The results of the proportions tests (the chi-squared values) are presented in Table 10. What these tests tell us, is whether there is a significant difference among the proportions of those, *per education category*, that answered "Yes" to a question about a simple speech compared to the "Yes"-answers in response to difficult speeches. If there is, we can determine that there is a difference of language complexity effect between different education levels.

Let us first look at the respondents holding an undergraduate degree, the largest group among the respondents. The proportions tests resulted in values of 3.25 (chi-squared) and 0.0714247 (p-value) for political support, and 2.6182 (chi-squared) and 0.1056454 (p-value) for political distance. This means that, while the political *distance* variable is not significant, the political *support* variable is. This means that there is a significant impact of having an undergraduate degree on the effect that language complexity has on a person's propensity to vote for a certain politician.

The same, however, cannot be said for those who went on to successfully pursue a postgraduate degree. The proportions test among this group resulted in very low chi-squared

Table 10: Proportions test results per education level

Category	*Political support: chi-squared*	*Political distance: chi-squared*
Undergraduate degree	3.25*	2.6182
Postgraduate degree	0	0.04137
Some university	3.2542*	1.356
Secondary school	0.10441	0.61562

** p<0.1 ** p<0.05 *** p<0.01*

scores and very high p-values, for both the political support and political distance variables. This means that the effect seen among the undergraduate degree holders is not repeated among those with a postgraduate degree.

The group who attended university but did not finish it, however, shows different results. While the values for political distance variable are not significant, the values for the political support variable are significant ($p<0.1$). This means that among this group, just like the undergraduate degree holders, there is a significant effect of the education level on the effect of language complexity on a person's propensity to vote for a certain politician.

Finally, those who listed secondary school as their highest completed level of education. The proportions tests for this group yielded no results for both political support and political distance, that would indicate an education effect on the effect of language complexity.

Bringing these results together, it becomes clear that the tests for education levels show different results than those for social classes did. Still, the effects are not overwhelming. For political *distance*, there is no significant effect of any education category influencing the effect of language complexity. However, among two groups (undergraduate degree holders and those with university experience but no diploma), the effect does exist for political *support*. This means that the effect of language complexity on a person's likelihood to vote for a politician (which is what the political support question asked the respondents) is *significantly impacted* by the education received by members of these two groups. Before moving on to analyzing the results that will help assess hypothesis 7, I will first show the results per respondents' party identification categories.

Differences per partisan identification categories

When analyzing respondents based on their partisan identification, it must be noted that there are multiple groups that can be collapsed into one. As noted above, a number of respondents can be classified as "non-partisans". I do so by taking all respondents who did not enter a specific party (e.g. Labour, Green Party) or ideological preference (such as "Central/somewhere in the middle"), and who did not refuse to answer. I also do not include the single respondent who answered "Other", as this person may have been referring to a party that was not listed and can therefore not be considered "non-partisan". Therefore, I create one analytical category of "non-partisans" by combining the respondents who gave the answers "A mixture/somewhere between various parties", "No/none", "Any party I agree with at the time", and "Independent". While the "A mixture" category may indicate that the person prefers a few parties over others, these respondents are not fixed on one specific party, which is why I feel it is acceptable to include them in this broader category. Together, the "non-partisans" make up 27.67% of the respondents, and are therefore the second largest group, only after the Labour supporters.

In the analysis here, I look at the respondents in six groups (Non-partisan, Labour, Liberal Democrat, Conservative, Green Party, and Central[14]). The results of the proportions tests (chi-squared values) are presented in Table 11. As before, these tests tell us whether there is a significant difference among the proportions of those, *per party identification category*, that answered "Yes" to a question about a simple speech compared to the "Yes"-answers in response

Table 11: Proportions test results per party identification

Category	*Political support: chi-squared*	*Political distance: chi-squared*
Non-partisan	0.48214	0
Labour	0.16746	0.090918
Liberal Democrat	1.9538	2.9727*
Conservative	1.1033	1.4903
Green Party	0.035885	0.30033
Central	1.5568	0.53684

** p<0.1 ** p<0.05 *** p<0.01*

[14] By Central I am referring to the group who answered "Central/somewhere in the middle".

to difficult speeches. If there is, we can determine that there is a difference of language complexity effect between people who identify with different parties.

For the non-partisans, Labour, Conservative, Green Party and Central supporters, no significant effects can be found for either the political support or political distance variables. This indicates that among the members of these groups, their party identification does not significantly impact the effect of language complexity.

The only group of partisans where a significant effect is visible, is the group of Liberal Democrat supporters. While the values for political support are not significant (chi-squared 1.9538, p-value 0.162), the values for political distance are significant (at the $p<0.1$ level). This means that identifying as a Liberal Democrat supporter has a significant effect on the effect that language complexity has on a person's likelihood to feel understood by a politician (which is what the political distance question asks the respondents).

So, does the party someone identifies with have an impact on the effect that language complexity will have on their propensity to support a politician? These results indicate no. Not one of the party groups shows a significant effect on the support variable. That said, with the political distance variable, the Liberal Democrat group does show an effect. So among those in the subject pool who identified with the Liberal Democrats, their party identification significantly impacted the effect that language complexity had on their feelings of being understood by the speaker.

Assessing influence of speaker party

However, while the above gives an image of how *respondents'* party identification conditions the effect of language complexity, to assess hypothesis 7 we need to look into how the *speaker's* party influences the effect of language complexity. In other words, if a Labour politician uses simple language in parliamentary speeches, does that have a significantly different effect compared to when a Conservative politician uses simple language? In order to investigate, I looked at what parties the speakers of the speeches presented to the respondents in this experiment represented. The dataset consisted of 8 speeches that had been selected (without regard for what party the speaker represented) from the pre-determined list of speeches on the debate topic of unemployment benefits, based solely on their Fleisch-Kincaid grade level score,

to determine language complexity. The selected speeches had been altered to either simple or difficult (depending on whether the original speech was simple or difficult), which doubled the number of speeches in the dataset to 16. In this final dataset, 8 speeches (4 simple, 4 difficult) were from Conservative politicians, 6 speeches (3 simple, 3 difficult) from Labour politicians, and 2 speeches (1 simple, 1 difficult) from Naomi Long, an MP for the Northern Irish APNI party. As APNI has not been present in the rest of the analysis, I exclude this party's influence on language complexity here, and focus solely on Labour and Conservative. The results of the proportions tests (chi-squared values) are presented in Table 12. What these results tell us is whether there is a significant difference between when a Labour politician uses simple language versus when a Conservative politician does so.

Table 12: Proportions tests comparing Labour and Conservative language complexity

Category	*Political support*	*Political distance*
Simple language	13.097***	22.546***
Difficult language	21.764***	14.666***

** p<0.1 ** p<0.05 *** p<0.01*

As can be seen in Table 12, the effects are highly significant for all four categories. When a Labour politician uses simple language, this has a significantly different effect on a person's propensity to vote for them than when a Conservative politician uses simple language. The same is true for difficult language, and the difference is similarly big for the political distance variable. Viewing the different proportions in Table 13, we can see that the proportions of people answering "Yes" on the political support (would vote for speaker) and political distance (feels understood by speaker) questions are higher when the speaker is representing the Labour Party. This holds for both simple and difficult speeches. This indicates that it matters, for the effect of language complexity which party the speaker represents. Remember that the respondents are not told what party the speaker represents, so the effect holds irrespective of

Table 13: Proportions of people answering "Yes" by speaker party

Category	*Political support: simple*	*Political support: difficult*	*Political distance: simple*	*Political distance: difficult*
Labour	0.6192469	0.5857741	0.5104603	0.4351464
Conservative	0.4626866	0.3850746	0.396462	0.2776119

what respondents think about a certain party or politician.

However, while there are significant differences between Labour and Conservative in the effects of the language complexity they use, the differences within the parties are less pronounced. In other words, when a Labour politician uses simple language, this does not have a significantly different effect on a respondent's likelihood to vote for them compared to when a Labour politician uses difficult language. This can be seen in Table 14, which displays the chi-squared values of the proportions tests where simple language by a party was compared to difficult language by that same party.

Table 14: Proportions tests comparing complexity per speaker party

Category	*Political support*	*Political distance*
Labour	0.42803	2.4256
Conservative	3.8199*	0.71903

** p<0.1 ** p<0.05 *** p<0.01*

With regards to Labour, the results indicate that when a Labour politician uses simpler language, this does not have a significant effect on a person's support for Labour, or their feeling of being understood by the Labour politician.

Meanwhile, for the Conservatives, the story is different. While we cannot say that when a Conservative politician uses simpler language, people will *feel more understood* by this politician. However, people will be more likely to *support* the Conservative politician. Given the significant difference of people indicating they would vote for a Conservative politician using simpler language versus a Conservative politician using difficult language, we can determine that when a Conservative politician uses simpler language, people will be more likely to vote for them. This is in line with the language recognition thesis (and the findings earlier in this chapter), which posits that when a politician uses simpler language, people will be more inclined to support them.

Influence of speaker party per social class

But while this refers to the speaker's party effect on *all* respondents, in order to assess hypothesis 7, we need to look at this effect per social class. Hypothesis 7 stated that by using less complex language, other parties can successfully appeal to the traditional support base of social democratic parties – the working class. What this would mean here, is that when a Conservative politician uses simpler language, this will have a significantly positive effect on their support among members of the working class.

To keep things clear, I limit the class categories here to 3: no class, working class, and middle class. First, let us look at the differences in proportions per social class for when Labour and Conservative politicians use simple language. Table 15 displays the results of the proportions tests (chi-square values).

Table 15: Labour and Conservative simple language per social class

Variable	*Political support*	*Political distance*
No class	3.411*	5.7639**
Working class	2.2788**	3.2645*
Middle class	5.3988**	8.4453***

** p<0.1 ** p<0.05 *** p<0.01*

What these results show, is that it matters what party is using simpler language. This shows a similar effect as before, when we investigated the speaker party effect for all voters. What this shows again is that people are significantly more likely to support a Labour politician using simpler language than a Conservative politician using simpler language. This effect holds across all the social classes analyzed. This, therefore, seems to indicate that hypothesis 7 is not supported.

However, these results do not necessarily mean that the Conservative politicians do not stand to benefit from using simple rather than difficult language to appeal to different social classes. Therefore, it is important to look at the differences between the proportions of people supporting Labour and the Conservatives when they use simple versus language per social class. Table 16 displays the results of this proportions test (chi-squared values).

When looking at the results, what becomes clear is that the parties have different effects for different social classes. The Conservative politicians stand to benefit most from using simpler

Table 16: Complexity per speaker party for different social classes

Variable	*Labour*	*Conservative*
No class: political support	0.21479	4.4504**
No class: political distance	0.375	2.4517
Working class: political support	0.18976	0.15068
Working class: political distance	0.046316	0
Middle class: political support	7.8072***	0.17469
Middle class: political distance	5.558**	0.19974

** p<0.1 ** p<0.05 *** p<0.01*

language among people who do not identify with a specific social class. People who do not identify with any social class are significantly more likely to support a Conservative politician when they use simpler language, versus when they use more difficult language in their speeches. This effect only holds for political support (the no class-respondents are not significantly more likely to feel understood by the Conservative politician). Moreover, the effect is not different for Labour politicians, meaning Labour does not stand to benefit from using simpler language among members of the no class category.

Where Labour does stand to benefit, is among those who identify as middle class. Both on the political support and political distance questions, middle class respondents are significantly more likely to answer "Yes" when a Labour politician uses simpler language versus when they use difficult language. This is an interesting finding and shows that Labour can attract more middle class supporters when they use less complex language.

The main group of interest for assessing hypothesis 7, however, is the working class. As the results clearly show, there is no significant effect (for both Labour and Conservative) of using simpler language versus more difficult language for members of the working class. What this means, is that even if the Conservative politicians use less complex language, working class respondents are not more likely to support them. This finding is surprising, as it is not in line with the expectations of the language recognition thesis, as this would make us expect to see a significant difference. That said, this result could have something to do with the topic of the speeches (unemployment benefits), which could be a topic where Labour has more "ownership"

(Budge, 2015), and therefore more support, irrespective of language complexity. This could also explain why the difference is also not present among working class individuals with regards to Labour language; it could be that people are more likely to support Labour's positions on this issue anyway. What these results mean, however, is that hypothesis 7 cannot be supported.

Conclusion and outlook

In this chapter, the results of the survey experiment conducted through the Nuffield Centre of Experimental Social Sciences (CESS) were discussed. After describing the characteristics of the respondents, a baseline proportions test was conducted, in order to see whether there would be a significant effect of language complexity on political support and political distance, as well as to determine whether the process of altering original parliamentary speeches delivered a confounding effect. The results were clear: people exposed to simpler parliamentary speeches were significantly more likely to both vote for the speaker (political support) and feel understood by the speaker (political distance). Furthermore, it was shown that there were no significant differences between responses for original and altered speeches, removing the concern of this being a potential confounder.

After this, a more fine-grained approach was taken, in which respondents' social characteristics were used to view the effects of language complexity. While no class effect on the effects of language complexity was found, the effects of undergraduate education (among both those who received a degree and those who did not) did yield significant results. Furthermore, respondents who identified as Liberal Democrat supporters were the only ones who displayed a significant party effect on language complexity's impact on the feeling of being understood by the speaker.

In order to assess hypothesis 7, the effects of what party the speaker represented were investigated. It was found that it matters what party the speaker represents. For both political support and political distance, people were significantly more likely to support the Labour Party than the Conservatives. However, people were found to be more likely to vote for Conservative politicians who use simpler language versus Conservative politicians who use difficult language.

When disaggregating the results, it became clear that across social classes, people remained more likely to support Labour politicians than Conservative politicians, when both use simpler language. That said, Conservative politicians do stand to benefit from increased support from people who do not identify with a social class, when they use less complex language in their speeches. The same is true for Labour, only among members of the middle class rather than those who do not identify with a class.

Crucially, however, it was found that among members of the working class, there was no significant difference in support levels when Conservative (or Labour) politicians used simpler versus more difficult language. This means that hypothesis 7 could not be supported by these results.

While these findings need to be read within the boundaries of the experiment (which only looked at parliamentary speeches on one debate topic), the findings nonetheless are remarkable and give important new insights. In the next chapter, the results of the regression chapter and this chapter will be discussed, after which conclusions will be drawn about the answer to the central research question of what explains the varying class-based support of social democratic parties.

Conclusion

What explains variations in the class-based support for social democratic parties? While demand-side explanations have been able to account for societal changes influencing individual-level characteristics that define what parties people were likely to support (i.e., substituting the traditional class cleavage by postmaterialist values), they have been unable to explain persistent class differences in support for social democratic parties. While supply-side explanations have been strong at explaining the role of party behavior and elite cues in (dis)incentivizing class voting, there have been doubts about these theories' explanatory power in certain time periods.

In order to contribute to this scholarship, I introduced the *language recognition thesis*, drawing on theories and methods from political science, sociology, linguistics and computer science. I argued that voters were looking for *recognition* when looking for representation, and that this could be fueled by politicians using *less complex language*. I argued that when the language used across parties becomes more complex, people will return to their traditional class identities to determine their vote choice.

The main findings in this thesis supported this argument. In the first step of my research, I investigated the majority of the hypotheses using a multilevel logistic regression model, using a pooled dataset of parliamentary speeches and survey data that covered thirty years of British politics. The analysis yielded support for some existing demand-side and supply-side explanations. Hypothesis 1, which posited that working class constituents since the 1990s have been less likely to support social democratic parties than middle class constituents, was supported by the findings. The same was true for hypothesis 2, which alleged that lower-educated constituents since the 1990s have been less likely to support social democratic parties than higher-educated constituents. Furthermore, hypothesis 3, which followed the supply-side political choice thesis in stating that more ideological polarization would lead to more class-based support for social democratic parties, was also supported by the findings.

Yet, the findings also suggested that when the language used in politics becomes *more* complex, support for social democratic parties will be *more* class-based. Hypothesis 4, which assumed that more complex language across parties would lead to more class-based support for social democratic parties was supported by the findings. The findings showed that when

controlling for polarization (hypothesis 3) and language complexity (hypothesis 4), individual-level characteristics became more significant and displayed more pronounced differences, with higher class groups becoming *less* inclined to support social democratic parties, while the working class became *more* inclined to support. Language complexity in politics, the findings pointed out, acts as a context that activates traditional class divisions.

The findings were not as supportive for hypothesis 5, which posited that educational differences in support for social democratic parties would increase when the language used in politics would become *more* complex. While people without formal qualifications were found to become *more* likely to support social democratic parties, all other groups were not turned away from social democratic parties. This meant that hypothesis 5 could not be confirmed.

With regards to debates on a topic "owned" by social democratic parties, the welfare state, hypothesis 6 assumed that *more* complex language across parties in debates on this topic would lead to *more* support for social democratic parties. The findings supported this, indicating that on the topic of the welfare state, social democratic parties benefit from an overall increase in language complexity across parties.

In the second step of my research, I used an original survey experiment to investigate the final hypothesis. Hypothesis 7 alleged that other parties that used *less* complex language would be able to successfully "lure" away working class constituents from social democratic parties. This assumption was not supported by the findings. However, another finding was in line with the language recognition thesis. Social democratic parties were shown to benefit from more support from middle class constituents when they used less complex language. This provides further support for the language recognition thesis' assumption that parties using less complex language can increase their support base beyond their traditional class-based constituency.

Implications of the findings

The findings in this thesis have implications for several strands of the literature. First, with regards to the literature on class voting. What the findings show is that there is room for a novel argument in explaining what factors increase or decrease the levels of class voting. The language recognition thesis provides such an argument. It adds important insights to our understanding of factors that influence class voting. When taking language complexity into

account, individual characteristics like social class become more important in determining party support. This means that theories advancing demand-side explanations for variations in class voting would benefit from adding the context of language complexity, in order to determine what this would mean for the impact of individual factors like social class, education (and age, gender and ethnicity). The results indicate that contextual factors like language complexity increase explanatory power. Furthermore, the results show that the context of language complexity impacts the effects of individual (class) characteristics on party support.

Second, with regards to the literature on the class-support basis for social democratic parties, the findings have important implications for the dilemma of what electorate social democratic parties should appeal to – the traditional core constituency of working-class voters, or the middle class constituents that have been appealed to since the 1990s (Kitschelt, 1994; Dalton, 2021). As noted earlier in this thesis, this dilemma has been discussed frequently (for instance in Przeworski, 1985; Kitschelt, 1994; Gingrich, 2017), with authors often noting that social democratic parties are well-served by appealing to the more culturally liberal yet economically moderate middle class (Kitschelt, 1994).

The findings here, however, show that the traditional working class base still matters for social democratic parties. In the context of more complex language use across parties, social democratic parties regain the support from the working class. However, what the experimental results showed is that individuals who identify as middle class are *more* likely to support social democratic parties when these parties use *less* complex language. This has important implications. When the language *across parties* becomes *more* complex, social democratic parties receive more support from their traditional base. However, when social democratic parties *themselves* use *less* complex language, they can successfully appeal to members of the middle class.

The findings in this thesis also have implications for literature on political communications. What the results indicate is that the perceived "recognition gap" among groups in society (Lamont, 2018) can be bridged by less complex language use in politics. Language complexity then can be viewed as a form of symbolic representation, where the representative increases the degree to which they are *believed in* by the represented (Pitkin, 1967: 102). This way, a politician who refrains from using "academic language" and instead will "[s]peak

normal" (Noordzij et al., 2010: 8) can increase how much people will believe in them as their representative. As the results presented in this thesis show, an example of this effect is less complex language use by social democratic parties leading to a more successful appeal to middle class voters.

Potential criticisms

Critics may claim that the relationship between parliamentary speeches and party support is not realistic, as they may believe most people do not follow parliamentary debates, and therefore would not be subjected to the language use in those debates. That is, however, not what I am suggesting. In this thesis, I have used parliamentary speeches as a proxy for all the communications by political parties, which also includes their communications in the media, through party manifestos, campaign speeches and social media as well. Parliamentary speeches offer a fine-grained data source that both politicians and parties view as a means to communicate to voters (Bäck et al., 2019: 165; Baumann et al., 2015; Benoit et al., 2012; Lauderdale and Herzog, 2016: 392; Poyet and Raunio, 2020: 20). How they communicate in parliamentary speeches therefore can be considered a good proxy for their overall communications.
Furthermore, in my experiment, I directly subjected participants to parliamentary speeches. The experimental findings supported the broader relationship suggested by the language recognition thesis: that recognition between the electorate and parties is increased when parties use less complex language.

Others may argue that my thesis implies that people who identify as working class and/or who have no formal qualifications are not smart enough to understand complex language used in parliament. This, however, would be an incorrect argument. I am not implying that there is a connection between someone being able to understand language used in parliament, and their social class or formal qualifications. Rather, my argument states that more difficult language use across parties will induce class-based voting for social democratic parties. This has nothing to do with intelligence, and the results support this. As my results in the British case show, working class individuals are *more* likely to support Labour when the language across parties becomes *more* complex, while individuals from the "salariat" group are *less* likely to support Labour under these conditions. Moreover, as my experiment shows, when Labour politicians use *less* complex language, people who identify is middle class are *more* likely to

support that politician. Therefore, implying that language complexity is tied to intelligence is not something I am suggesting nor something supported by the findings.

Future work

While the findings in this thesis are promising, there are some avenues that future work could explore. First is the use of more fine-grained data sources. An unavoidable limitation I encountered when conducting the research for this thesis was the low frequency of surveys that were freely available. This was particularly disappointing when considering that the language data used provided almost daily data (parliamentary speeches). Because of the lack of a more frequent data source, I had to use yearly survey data from the British Social Attitudes survey, supplemented with even less frequent survey data from the British Election Study. A solution to this problem could have been using (internet) panel study data, such as the panel data offered by the British Election Study. Unfortunately, however, this data contains gaps, for instance between 2001 and 2005, as well as 2010 and 2014 (British Election Study, 2021b). This meant that I had to either use a shorter time frame (one covered by the internet panel data) or the longer time frame (covered by the yearly survey data). In this study, I opted for the latter, studying the yearly data between 1989 and 2019. Future work could use (internet) panel data to get a more fine-grained data source in those shorter time periods. Surveys that consistently produce more regular results on the same variables, however, would also be hugely beneficial to this type of work.

Another avenue that future work could focus on is expanding the scope of the topics investigated in the experimental setting. Because of financial constraints, the experimental research for this thesis was unable to include a wide range of topics, to see whether there would be a topic-related effect on language complexity in parliamentary speeches. As the regression results showed, such a topic effect can be found for a topic "owned" by social democratic parties (the welfare state). Future studies could add important insights through experiments using a wide range of topics, either as a control on the relationship between language complexity and party support or to investigate under what topics less complex language has the greatest effect on support. In this way, future work could also investigate under what topics which parties are more likely to receive more support by using less complex language.

In this thesis, however, I introduced the language recognition thesis as a novel argument, linking the complexity of language political parties use to class-based support for social democratic parties. I used methods from other disciplines, like linguistics and computer science, to analyze the complexity of thirty years of parliamentary speeches from the UK House of Commons. I then investigated my hypotheses using two research methods: a multilevel logistic regression and a survey experiment. The findings supported my expectations that when the language used across political parties becomes *more* complex, there will be more class-based support for social democratic parties.

References

Abou-Chadi, T. and J. Gingrich (2021) "It's not just in Britain across Europe, social democracy is losing its way" *The Guardian*, 9 May 2021 < https://www.theguardian.com/commentisfree/2021/may/09/not-just-britain-across-europe-social-democracy-losing-way> (Accessed 15 May 2021)

Abou-Chadi, T. and M. Wagner (2020) "Electoral fortunes of social democratic parties: do second dimension positions matter?" *Journal of European Public Policy* 27(2), 246-272

Abramson, P.R. and R. Inglehart (1992) "Generational Replacement and Value Change in Eight West European Societies" *British Journal of Political Science* 22(2), 183-228

Advance HE (2020) "Social background - Guidance on equality monitoring in relation to social background." <advance-he.ac.uk/guidance/equality-diversity-and-inclusion/using-data-and-evidence/monitoring-questions/social-background> (Accessed 12 July 2021)

Arzheimer, K. and E. Carter (2006) "Political opportunity structures and right-wing extremist party success" *European Journal of Political Research* 45(3), 419-443

Arzheimer, K. (2009) "Contextual Factors and the Extreme Right Vote in Western Europe, 1980–2002" *American Journal of Political Science* 53(2), 259-275

Bäck, H., M. Baumann, M. Debus and J. Müller (2019) "The Unequal Distribution of Speaking Time in Parliamentary-Party Groups" *Legislative Studies Quarterly* 44(1), 163-193

Bakker, R. and S. Hobolt (2013) "Measuring Party Positions", in: in: G. Evans and N.D. de Graaf (eds.) *Political Choice Matters: Explaining the Strength of Class and Religious Cleavages in Cross-National Perspective* (Oxford: Oxford University Press), 27-45

Baumann, M., M. Debus and J. Müller (2015) "Convictions and Signals in Parliamentary Speeches: Dáil Éireann Debates on Abortion in 2001 and 2013" *Irish Political Studies* 30(2), 199-219

Bawn, K., M. Cohen, D. Karol, S. Masket, H. Noel and J. Zaller (2012) "A Theory of Political Parties: Groups, Policy Demands and Nominations in American Politics" *Perspectives on Politics* 10(3), 571-597

Beackon, S. (1976) "Labour Party Politics and the Working Class" *British Journal of Political Science* 6(2), 231-238

Benoit, K., K. Munger and A. Spirling (2019) "Measuring and Explaining Political Sophistication through Textual Complexity" *American Journal of Political Science* 63(2), 491-508

Benoit, K., D. Schwarz and D. Traber (2012) "The Sincerity of Political Speech in Parliamentary Systems: A Comparison of Ideal Points Scaling Using Legislative Speech and Votes" (Conference Paper)

Best, R.E. (2011) "The declining electoral relevance of traditional cleavage groups" *European Political Science Review* 3(2), 279-300

Bischof, D. and R. Senninger (2018) "Simple politics for the people? Complexity in campaign messages and political knowledge" *European Journal of Political Research* 57(2), 473-495

Blei, D.M., A.Y. Ng and M.I. Jordan (2003) "Latent Dirichlet Allocation" *Journal of Machine Learning Research* 3, 993-1022

Blei, D.M. (2012) "Probabilistic Topic Models" *Communications of the ACM* 55(4), 77-84

British Election Study (2021a) "Cross-sectional data" <https://www.britishelectionstudy.com/data-objects/cross-sectional-data/> (Accessed 24 May 2021)

British Election Study (2021b) “Panel study data” <https://www.britishelectionstudy.com/data-objects/panel-study-data/> (Accessed 09 October 2021)
British Social Attitudes (2014) *BRITISH SOCIAL ATTITUDES 2014 DOCUMENTATION OF THE BLAISE QUESTIONNAIRE* (UK Data Archive Study Number 7809)
Brook, L., C. Chaudhary, A. Park and K. Thomson (1994) *British Social Attitudes: Young People's Social Attitudes and Northern Ireland Social Attitudes 1994 Surveys Technical Report* (Social & Community Planning Research)
Budge, I. (2015) “Issue Emphases, Saliency Theory and Issue Ownership: A Historical and Conceptual Analysis” *West European Politics* 38(4), 761-777
Burst, T., W. Krause, P. Lehmann, J. Lewandowski, T. Matthieß, N. Merz, S. Regel and L. Zehnter (2020) *Manifesto Corpus*, Version: 2020-2 (Berlin: WZB Berlin Social Science Center)
Butler, D. and D. Stokes (1977) *Political Change in Britain: The Evolution of Electoral Choice* Second Edition – Paperback Version (London: The Macmillan Press Ltd)
Cann, D.M., G. Goelzhauser and K. Johnson (2014) “Analyzing Text Complexity in Political Science Research” *PS: Political Science & Politics* 47(3), 663-666
Carnes, N. and N. Lupu (2016) “Do Voters Dislike Working-Class Candidates? Voter Biases and the Descriptive Underrepresentation of the Working Class” *American Political Science Review* 110(4), 832-844
Chamberlain, C.W. and H.F. Moorhouse (1974) “Lower Class Attitudes towards the British Political System” *The Sociological Review* 22(4), 503–526
Clark, C. (2021) “BBC's Chris Mason defends interview with angry ex-Labour voters after left-wingers erupt” <https://www.express.co.uk/news/politics/1434108/bbc-chris-mason-interview-election-labour-voters-video> (Accessed 04 October 2021)
Clark, T.N. and S.M. Lipset (1991) “Are Social Classes Dying?” *International Sociology* 6(4), 397-410
Clarke, H.D., D. Sanders, M.C. Stewart and P. Whiteley (2004) *Political Choice in Britain* (Oxford: Oxford University Press)
Crabtree, C., M. Golder, T. Gschwend and I.H. Indriđason (2020) “It Is Not Only What You Say, It Is Also How You Say It: The Strategic Use of Campaign Sentiment” *The Journal of Politics* 82(3), 1044-1060
Curran, B., K. Higham, E. Ortiz and D. Vasques Filho (2018) “Look who’s talking: Two-mode networks as representations of a topic model of New Zealand parliamentary speeches” *PLoS ONE* 13(6), 1-16
Dalton, R.J. (2021) “Modeling ideological polarization in democratic party systems” *Electoral Studies* 72, p1-10
Dalton, R.J. (2008) “The Quantity and the Quality of Party Systems: Party System Polarization, Its Measurement, and Its Consequences” *Comparative Political Studies* 41(7), 899-920
Decadri, S. and C. Boussalis (2020) “Populism, party membership, and language complexity in the Italian chamber of deputies” *Journal of Elections, Public Opinion and Parties* 30(4), 484-503
Douma, W.H. (1960) *De Leesbaarheid van Landbouwbladen: Een Onderzoek naar en Toepassing van Leesbaarheidsformules* (Wageningen: Landbouwhogeschool Wageningen)
Eggers, A.C. and A. Spirling (2017) “Incumbency Effects and the Strength of Party Preferences: Evidence from Multiparty Elections in the United Kingdom” *The Journal of Politics* 79(3), 903-920
Esping-Andersen, G. (1985) *Politics Against Markets: The Social Democratic Road to Power*

(Princeton: Princeton University Press)
Evans, G. and J. Mellon (2016) *Social Class – Identity, awareness and political attitudes: why are we still working class?* (London: NatCen Social Research) <https://www.bsa.natcen.ac.uk/media/39094/bsa33_social-class_v5.pdf> (Accessed 30 September 2021)
Evans, G. and J. Tilley (2017a) *The new politics of class in Britain: The political exclusion of the working class* (Oxford: Oxford University Press)
Evans, G. and J. Tilley (2017b) "The New Politics of Class after the 2017 General Election" *The Political Quarterly* 88(4), 710-715)
Evans, G. and J. Tilley (2013) "Ideological Convergence and the Decline of Class Voting in Britain", in: G. Evans and N.D. de Graaf (eds.) *Political Choice Matters: Explaining the Strength of Class and Religious Cleavages in Cross-National Perspective* (Oxford: Oxford University Press), 87-113
Evans, G. (2000) "The Continued Significance of Class Voting" *Annual Review of Political Science* 3, 401-417
Flesch, R. (1948) "A New Readability Yardstick" *Journal of Applied Psychology* 32(3), 221-233
Ford, R. and W. Jennings (2020) "The Changing Cleavage Politics of Western Europe" *Annual Review of Political Science* 23, 295-314
Gingrich, J. (2017) "A New Progressive Coalition? The European Left in a Time of Change" *The Political Quarterly* 88(1), 39-51
Gingrich, J. and S. Häusermann (2015) "The decline of the working-class vote, the reconfiguration of the welfare support coalition and consequences for the welfare state" *Journal of European Social Policy* 25(1), 50-75
Goldthorpe, J.H., D. Lockwood, F. Bechhofer and J. Platt (1968) *The Affluent Worker: Political attitudes and behaviour* (Cambridge: Cambridge University Press)
Goldthorpe, J.H (2016) "Social class mobility in modern Britain: changing structure, constant process" *Journal of the British Academy* 4, p89-111
Greene, D. and J.P. Cross (2017) "Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach" *Political Analysis* 25(1), 77-94
Heath, A., I. McLean, B. Taylor and J. Curtice (1999) "Between first and second order: a comparison of voting behaviour in European and local elections in Britain" *European Journal of Political Research* 35(3), 389-414
Heath, O. (2015) "Policy Representation, Social Representation and Class Voting in Britain" *British Journal of Political Science* 45(1), 173-193
Hindess, B. (1971) *The Decline of Working-Class Politics* (London: Paladin)
Hlavac, M. (2018) *stargazer: Well-Formatted Regression and Summary Statistics Tables*, R package version 5.2.2. <https://CRAN.R-project.org/package=stargazer>
Houtman, D. (2003) "Lipset and "Working-Class" Authoritarianism" *The American Sociologist* 34, 85-103
Hox, J.J., M. Moerbeek and R. van der Schoot (2018) *Multilevel Analysis – Techniques and Applications,* Third Edition (New York: Routledge)
Inglehart, R. and P.R. Abramson (1994) "Economic Security and Value Change" *The American Political Science Review* 88(2), 336-354
Inglehart, R. and J-R. Rabier (1986) "Political Realignment in Advanced Industrial Society: From Class-Based Politics to Quality-of-Life Politics" *Government and Opposition* 21(4), 456-479
Inglehart, R. (1977) *The Silent Revolution: Changing Values and Political Styles Among Western Publics* (Princeton, NJ: Princeton University Press)
Jansen, D-J. (2011) "Does The Clarity Of Central Bank Communication Affect Volatility In

Financial Markets? Evidence From Humphrey-Hawkins Testimonies" *Contemporary Economic Policy* 29(4), 494-509
Jansen, G., G. Evans and N.D. de Graaf (2013) "Class Voting and Left-Right Party Positions", in: G. Evans and N.D. de Graaf (eds.) *Political Choice Matters: Explaining the Strength of Class and Religious Cleavages in Cross-National Perspective* (Oxford: Oxford University Press), 46-82
Karon, B.P. (1964) "A Note on the Treatment of Age as a Variable in Regression Equations" *The American Statistician* 18(3), 27-28
Kincaid, J.P., R.P. Fishburne Jr., R.L. Rogers and B.S. Chissom (1975) *Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel* (Millington, Tennessee: Chief of Naval Technical Training Naval Air Station Memphis)
Kitschelt, H. (1994) *The Transformation of European Social Democracy* (Cambridge: Cambridge University Press
Kosmidis, S., S.B. Hobolt, E. Molloy and S. Whitefield (2019) "Party Competition and Emotive Rhetoric" *Comparative Political Studies* 52(6), 811-837
Lamont, M. (2018) "Addressing Recognition Gaps: Destigmatization and the Reduction of Inequality" *American Sociological Review* 83(3), 419-444
Lauderdale, B.E. and A. Herzog (2016) "Measuring Political Positions from Legislative Speech" *Political Analysis* 24(3), 374-394
Laver, M.J. and I. Budge (1992) *Party Policy and Government Coalitions* (New York: St Martin's Press)
Lazarsfeld, P.F., B. Berelson and H. Gaudet (2021) *The People's Choice – How the voter makes up his mind in a presidential campaign* (New York: Columbia University Press)
Lee, L. and P. Young (2013) *Politics: A disengaged Britain? Political interest and participation over 30 years* (London: NatCen Social Research) <https://www.bsa.natcen.ac.uk/media/38455/bsa30_politics_final.pdf> (Accessed 30 September 2021)
Lefkofridi, Z. M. Wagner and J.E. Willman (2014) "Left-Authoritarians and Policy Representation in Western Europe: Electoral Choice across Ideological Dimensions" *West European Politics* 37(1), 65-90
Lewandowski, J., N. Merz and S. Regel (2020) *manifestoR: Access and Process Data and Documents of the Manifesto Project*, R package version 1.5.0 <https://CRAN.R-project.org/package=manifestoR>
Lijphart, A. (2012) *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries,* 2nd edition (New Haven: Yale University Press)
Lin, N. and M. Osnabrügge (2018) "Making comprehensible speeches when your constituents need it" *Research and Politics* 5(3), 1-8
Lipset, S.M. (1959) "Democracy and Working-Class Authoritarianism" *American Sociological Review* 24, 482-502
Lipset, S.M. and S. Rokkan (1967) *Party systems and voter alignments: cross-national perspectives* (New York: The Free Press)
Lüdecke, D. (2021) *sjPlot: Data visualization for statistics in social science* (Version 2.8.9) [R package]. (Comprehensive R Archive Network) https://CRAN.R-ssproject.org/package=sjPlot
Manza, J., M. Hout and C. Brooks (1995) "Class Voting in Capitalist Democracies Since World War II: Dealignment, Realignment, or Trendless Fluctuation?" *Annual Review of Sociology* 21(1), 137-162
Marcinkiewicz, K. & M. Stegmaier (2019) "Speaking up to stay in parliament: the electoral

importance of speeches and other parliamentary activities" *The Journal of Legislative Studies* 25(4), 576-596
Marx, P. and G. Schumacher (2018) "Do Poor Citizens Vote for Redistribution, Against Immigration or Against the Establishment? A Conjoint Experiment in Denmark" *Scandinavian Political Studies* 41(3), 263-282
McKenzie, D. (2012) "A pre-analysis plan checklist", World Bank Blogs <https://blogs.worldbank.org/impactevaluations/a-pre-analysis-plan-checklist> (Accessed 7 July 2021)
NatCen Social Research (2017) *British Social Attitudes 2017 User guide* (London: NatCen Social Research)
Noordzij, K., W. de Koster and J. van der Waal (2020) ""They don't know what it's like to be at the bottom": Exploring the role of perceived cultural distance in less-educated citizens' discontent with politicians" *British Journal of Sociology* 00, 1-14
Nosek, B.A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson and T. Yarkoni (2012) "Promoting an open research culture" *Science* 348(6242), 1422-1425
OECD (2021) "United Kingdom – Overview of the education system (EAG 2021)" <https://gpseducation.oecd.org/CountryProfile?primaryCountry=GBR&treshold=10&topic=EO> (Accessed 30 September 2021)
Oesch, D. (2008a) "The Changing Shape of Class Voting" *European Societies* 10(3), 329-355
Oesch, D. (2008b) "Explaining Workers' Support for Right-Wing Populist Parties in Western Europe: Evidence from Austria, Belgium, France, Norway, and Switzerland" *International Political Science Review* 29(3), 349-373
Orriols, L. (2013) "Social Class, Religiosity, and Vote Choice in Spain, 1979–2008", in: G. Evans and N.D. de Graaf (eds.) *Political Choice Matters: Explaining the Strength of Class and Religious Cleavages in Cross-National Perspective* (Oxford: Oxford University Press), 360-387
Peskowitz, Z. (2017) "Ideological Signaling and Incumbency Advantage" *British Journal of Political Science* 49(2), 467-490
Pitkin, H.F. (1967) *The Concept of Representation* (Berkeley: University of California Press)
Poyet, C. and T. Raunio (2020) "Reconsidering the Electoral Connection of Speeches: The Impact of Electoral Vulnerability on Legislative Speechmaking in a Preferential Voting System" *Legislative Studies Quarterly* 0(0), 1-26
Przeworski, A. (1985) *Capitalism and Social Democracy* (Cambridge: Cambridge University Press)
Przeworski, A. (1980) "Social Democracy as a Historical Phenomenon" *New Left Review* I/122, 27-58
Proksch, S-O and J.B. Slapin (2012) "Institutional Foundations of Legislative Speech" *American Journal of Political Science* 56(3), 520-537
Pulzer, P. (1987) "The Paralysis of the Centre-Left: A Comparative Perspective" *The Political Quarterly* 58(4), 378-388
Pulzer, P. (1967) *Political Representation and Elections in Britain* (London: George Allen and Unwin Ltd)
Rauh, C. and J. Schwalbach (2020) "The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies", <https://doi.org/10.7910/DVN/L4OAKN>, Harvard Dataverse, V1

Rennwald, L. and G. Evans (2014) "When Supply Creates Demand: Social Democratic Party Strategies and the Evolution of Class Voting" *West European Politics* 37(5), 1108-1135
Robison, J., R. Stubager, M. Thau and J. Tilley (2021) "Does Class-Based Campaigning Work? How Working Class Appeals Attract and Polarize Voters" *Comparative Political Studies* 54(5), 723-752
Schoonvelde, M., A. Brosius and B.N. Bakker (2019) "Liberals lecture, conservatives communicate: Analyzing complexity and ideology in 381,609 political speeches" *PLoS ONE* 14(2), 1-15
Seawright, J. and J. Gerring (2008) "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options" *Political Research Quarterly* 61(2), 294-308
Skinner, G., K. Pedley and D. Spielman (2019) *How the voters voted in the 2019 election* (London: Ipsos MORI)
Spirling, A. (2016) "Democratization and Linguistic Complexity: The Effect of Franchise Extension on Parliamentary Discourse, 1832–1915" *The Journal of Politics* 78(1), 120-136
Steinmetz, G. and E.O. Wright (1989) "The Fall and Rise of the Petty Bourgeoisie: Changing Patterns of Self-Employment in the Postwar United States" *American Journal of Sociology* 94(5), 973-1018
Tilley, J., J. Garry and N. Matthews (2021) "The Evolution of Party Policy and Cleavage Voting under Power-Sharing in Northern Ireland" *Government and Opposition* 56(2), 226-244
Tilley, J. and G. Evans (2014) "Ageing and generational effects on vote choice: Combining cross-sectional and panel data to estimate APC effects" *Electoral Studies* 33, p19-27
UK Data Service (2021) "British Social Attitudes Survey" <https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=200006#!/access-data> (Accessed 24 May 2021)
UK Government (2021a) "Departments, agencies and public bodies" <https://www.gov.uk/government/organisations> (Accessed 19 September 2021)
UK Government (2021b) "List of ethnic groups" <https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups> (Accessed 19 September 2021)
Van der Waal, J., P. Achterberg and D. Houtman (2007) "Class Is Not Dead—It Has Been Buried Alive: Class Voting and Cultural Voting in Postwar Western Societies (1956–1990)" *Politics & Society* 35(3), 403-426
Valentim, V. and T. Widmann (2021) "Does Radical-Right Success Make the Political Debate More Negative? Evidence from Emotional Rhetoric in German State Parliaments" *Political Behavior*, DOI: https://doi.org/10.1007/s11109-021-09697-8
Volkens, A., T. Burst, W. Krause, P. Lehmann, T. Matthieß, N. Merz, S. Regel, B. Weßels and L. Zehnter (2020) *The Manifesto Data Collection. Manifesto Project (MRG / CMP / MARPOR)*, Version 2020b (Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB)) <https://doi.org/10.25522/manifesto. mpds.2020b>
Wang, Y. and H. Liu (2018) "Is Trump always rambling like a fourth-grade student? An analysis of stylistic features of Donald Trump's political discourse during the 2016 election" *Discourse & Society* 29(3), 299-323
Waugh, E. (2000) *Brideshead Revisited: The Sacred and Profane Memories of Captain Charles Ryder* (London: Penguin)
Weakliem, D.L. (2002) "The Effects of Education on Political Opinions: An International Study" *International Journal of Public Opinion Research* 13(2), 141-157
Weakliem, D.L. (1991) "The Two Lefts? Occupation and Party Choice in France, Italy, and

the Netherlands" *American Journal of Sociology* 96(6), 1327-1361
Wickham-Jones, M. (1996) *Economic Strategy and the Labour Party: Politics and policy-making, 1970–83* (London: Palgrave Macmillan)

Appendices A-F

Appendix A – Examples of parliamentary speeches with complexity scores

Speech with a very high complexity (Flesch-Kincaid Grade Score level 57.4), delivered by Bob Seely MP (Conservative) on 22-07-2019 during a debate on the Telecoms Supply Chain Review:

> *"Given that Huawei is to all intents part of the Chinese state, given that China has extensive history of intellectual property theft, data theft, cyber-attacks and the development of a surveillance state in parts of its own country, given that it is building up a dominant position in advanced comms that will eventually put Nokia, Samsung and others out of business, given the increased warnings of the Cell, and given our Five Eyes colleagues' positions, does the Secretary of State agree that having any tech from one-party authoritarian states in our critical national infrastructure raises difficult and potentially insurmountable obstacles when it comes to data protection and protecting our human rights, the rule of law, our value system in the 21st century and security?"*

As can be seen, this text consists of one really long sentence, consisting of 124 words, which consist of 209 syllables. This makes the text quite difficult.

Speech with low complexity, spoken by John Hayes MP (Conservative) on 24-03-1998, during a debate on Home School Agreements (Flesch-Kincaid Grade Level score of 8.3):

> *"My hon. Friend is right to say that the Secretary of State has those powers. The hon. Member for Harrogate and Knaresborough (Mr. Willis) fails to appreciate that my hon. Friend is pointing out the critical and implicit difference between a contract, a partnership and a policy, which lies at the heart of the amendment moved by the hon. Member for Bath (Mr. Foster). The Government started off with a contract, we are now talking about an agreement, and the hon. Member for Bath is talking about a policy. Each is different, and my hon. Friend is drawing attention to that in an appropriate and entirely apposite way."*

As can be seen here, the 9 sentences used here are shorter (mean of 12 words per sentence), and the 108 words consist of fewer syllables.

Appendix B – Debate topics with their 15 most frequent words

THE TOP 15 WORDS FOR TOPIC #0: Foreign Affairs

['britain', 'british', 'international', 'hon', 'states', 'world', 'country', 'united', 'trade', 'europe', 'uk', 'countries', 'eu', 'union', 'european']

THE TOP 15 WORDS FOR TOPIC #1: General debate

['parliament', 'motion', 'deputy', 'friend', 'order', 'committee', 'point', 'time', 'speaker', 'mr', 'members', 'debate', 'member', 'house', 'hon']

THE TOP 15 WORDS FOR TOPIC #2: Welfare state

['money', 'make', 'government', 'pension', 'friend', 'women', 'young', 'need', 'benefit', 'help', 'scheme', 'support', 'hon', 'work', 'people']

THE TOP 15 WORDS FOR TOPIC #3: Housing

['council', 'private', 'register', 'houses', 'property', '000', 'want', 'local', 'does', 'home', 'problem', 'homes', 'hon', 'housing', 'people']

THE TOP 15 WORDS FOR TOPIC #4: General debate 2

['just', 'minister', 'did', 'say', 'member', 'state', 'secretary', 'said', 'members', 'conservative', 'hon', 'government', 'people', 'party', 'labour']

THE TOP 15 WORDS FOR TOPIC #5: Social care/services

['parents', 'nhs', 'service', 'right', 'funding', 'local', 'services', 'friend', 'school', 'care', 'hon', 'children', 'schools', 'education', 'health']

THE TOP 15 WORDS FOR TOPIC #6: General debate 3

['time', 'day', 'year', 'debate', 'report', 'prime', 'week', 'post', 'statement', 'government', 'office', 'house', 'minister', 'state', 'secretary']

THE TOP 15 WORDS FOR TOPIC #7: Finance

['rate', 'gentleman', 'spending', '000', 'years', 'money', 'billion', 'budget', 'increase', 'hon', 'million', 'year', 'cent', 'tax', 'government']

THE TOP 15 WORDS FOR TOPIC #8: Infrastructure

['does', 'areas', 'water', 'road', 'east', 'rail', 'constituents', 'member', 'south', 'north', 'constituency', 'right', 'west', 'friend', 'hon']

THE TOP 15 WORDS FOR TOPIC #9: Justice

['courts', 'law', 'friend', 'right', 'prison', 'court', 'criminal', 'people', 'cases', 'officers', 'justice', 'hon', 'home', 'crime', 'police']

THE TOP 15 WORDS FOR TOPIC #10: General debate 4

['report', 'evidence', 'question', 'house', 'said', 'point', 'case', 'matter', 'committee', 'right', 'gentleman', 'mr', 'friend', 'member', 'hon']

THE TOP 15 WORDS FOR TOPIC #11: Health care

['provision', 'minister', 'provisions', 'service', 'information', 'hon', 'patients', 'amendments', 'legislation', 'nhs', 'health', 'act', 'amendment', 'new', 'clause']

THE TOP 15 WORDS FOR TOPIC #12: Defense

['support', 'war', 'foreign', 'international', 'military', 'government', 'iraq', 'armed', 'security', 'friend', 'united', 'right', 'forces', 'hon', 'defence']

THE TOP 15 WORDS FOR TOPIC #13: Home/domestic affairs

['political', 'community', 'law', 'does', 'women', 'world', 'british', 'hon', 'prime', 'government', 'country', 'minister', 'human', 'people', 'rights']

THE TOP 15 WORDS FOR TOPIC #14: Transportation and Enegery

['communities', 'need', 'transport', 'authority', 'council', 'development', 'planning', 'government', 'new', 'friend', 'energy', 'authorities', 'london', 'hon', 'local']

THE TOP 15 WORDS FOR TOPIC #15: Fisheries and Farming

['animals', 'sea', 'country', 'british', 'fishing', 'agriculture', 'ban', 'friend', 'does', 'research', 'minister', 'hon', 'farmers', 'industry', 'food']

THE TOP 15 WORDS FOR TOPIC #16: Devolution-related/state-level affairs

['state', 'government', 'parliament', 'england', 'hon', 'people', 'assembly', 'welsh', 'kingdom', 'united', 'wales', 'scottish', 'scotland', 'northern', 'ireland']

THE TOP 15 WORDS FOR TOPIC #17: General debate 5

['paper', 'way', 'issues', 'public', 'forward', 'commission', 'legislation', 'consultation', 'members', 'report', 'proposals', 'house', 'hon', 'committee', 'government']

THE TOP 15 WORDS FOR TOPIC #18: General debate 6

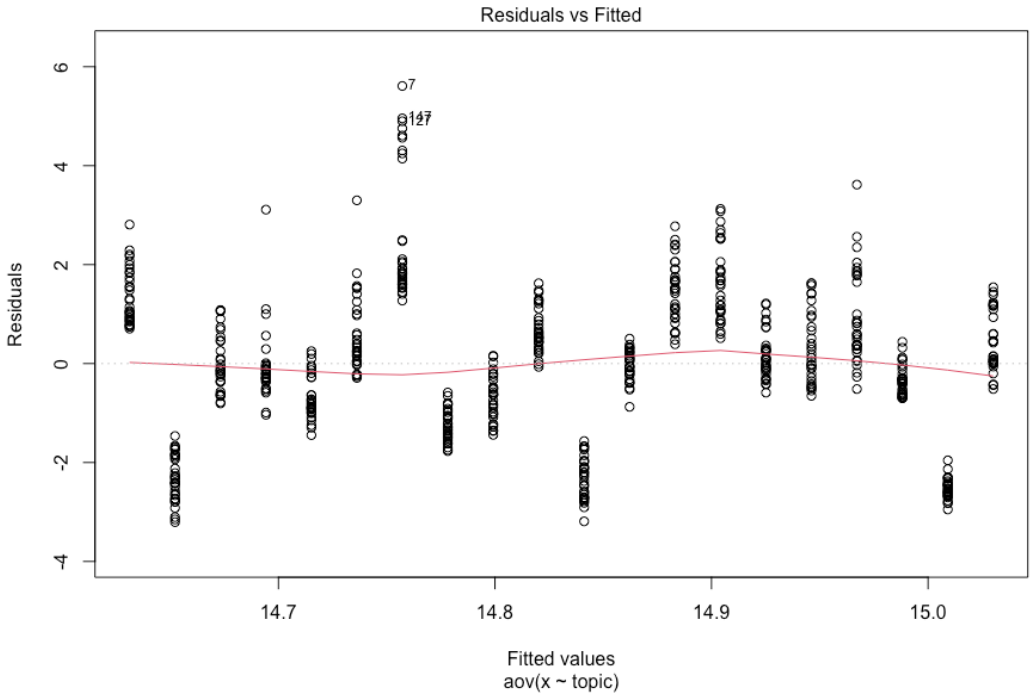
['agree', 'people', 'way', 'point', 'think', 'say', 'house', 'said', 'does', 'minister', 'government', 'friend', 'gentleman', 'right', 'hon']

THE TOP 15 WORDS FOR TOPIC #19: Economy

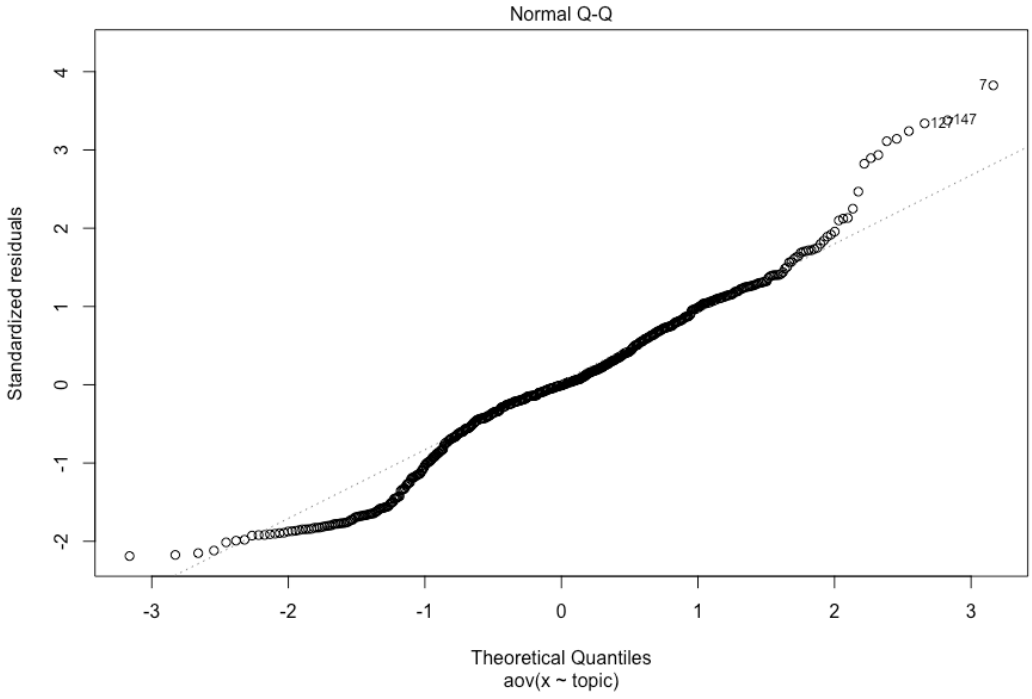
['market', 'economy', 'private', 'investment', 'company', 'british', 'small', 'hon', 'government', 'jobs', 'sector', 'businesses', 'companies', 'business', 'industry']

Appendix C – ANOVA Figures for the topic effects

Homogeneity of variances: no clear relationship between residuals and fitted values



Normality assumption: residuals normally distributed



Appendix D – Full version of Table 3

Table 3: individual-level model and full model

Dependent variable: Labour support

	Individual-level only model	**Full model**
Predictors	*Odds Ratios*	*Odds Ratios*
(Intercept)	0.00 ***	0.00
Foremen or Technicians	1.39 ***	1.16
Petty Bourgeois	1.16 **	0.64 **
Salariat (Higher & Lower)	0.81 ***	0.80 **
Working class	1.60 ***	1.41 ***
Below GCSE	1.19 ***	1.03
GCSE	1.10 *	1.09
No qualifications	1.53 ***	1.49 ***
Postgraduate	1.30 ***	1.09
Undergraduate	1.16 ***	0.96
Age	0.99 ***	0.81 ***
Log Age	1.33	1.32
Male	0.98	0.96
Asian or Asian British	2.64 ***	1.71
Black, African, Caribbean or Black British	3.33 ***	2.34 **
Mixed or Multiple ethnic groups	1.87 ***	0.93
White	1.16	0.72
Labour government		3.47 **
Party system polarization		1.97
Complexity		0.56
Welfare state complexity		2.75 ***
Random Effects		
σ^2	3.29	3.29

τ_{00}	55.81 survey_year	2.54 survey_year
τ_{11}		0.09 survey_year.socialclassForemen/Technicians
		0.37 survey_year.socialclassPetty Bourgeois
		0.04 survey_year.socialclassSalariat (Higher & Lower)
		0.12 survey_year.socialclassWorking class
		0.05 survey_year.edlevelBelow GCSE
		0.00 survey_year.edlevelGCSE
		0.03 survey_year.edlevelNo qualifications
		0.01 survey_year.edlevelPostgrad
		0.03 survey_year.edlevelUndergraduate
		0.05 survey_year.age
		0.01 survey_year.sex
		0.51 survey_year.ethnicityAsian or Asian British
		0.68 survey_year.ethnicityBlack, African, Caribbean or Black British
		0.49 survey_year.ethnicityMixed or Multiple ethnic groups
		0.73 survey_year.ethnicityWhite
ρ_{01}		0.23
		0.19
		-0.49
		-0.10
		0.64
		0.48
		-0.58
		0.86
		0.65
		0.60
		0.41
		-0.54
		-0.43
		-0.14
		-0.48
ICC	0.94	
N	29 survey_year	29 survey_year
Observations	77605	77605
Marginal R^2 / Conditional R^2	0.003 / 0.944	0.803 / NA
Pseudo R^2	0.00	0.07

** p<0.1 ** p<0.05 *** p<0.01*

Appendix E – Pre-analysis Plan

According to the principles of the Transparency and Openness Promotion (TOP) Guidelines, it is considered good research practice to describe a pre-analysis plan prior to conducting research (Nosek et al., 2015: 1423). Below relevant aspects of the study are described prior to analysis taking place. The below bullet points are taken from the suggested outline of a pre-analysis plan described on the website of the World Bank (McKenzie, 2012).

A. Sample and randomization

The sample used in this experiment will have the size of N = 333. This is derived from a combination of the financial constraints to the study (placed by the available funding and desire to provide reasonable payment to all participants) and the sample size calculation based on the desired effect size. Using a two-sampled t-test, and using a statistical power of 0.80 (80%) and significance level of 5%, the desired effect size was taken from an article with a similar research topic to this project (Robison et al., 2021). The sample size calculation yielded a required sample size of at least N=16 per testing group.
As my design includes two factors that take on two values, this implies a minimum total sample size of 64 to be able to detect my desired effect size. The increased sample size of 333 (about 83 per group) will allow me to do sub-group analysis based on the survey responses, for instance on social characteristics, which are asked after the experiment.

The subject pool will be randomly drawn from the CESS Subject Pool, with the only requirements that the participants need to be eligible voters in Great Britain, as the experiment looks into voting behavior in this country. Subjects will be randomly assigned to the two treatments. Using a within-subjects design, all subjects will receive all treatments and controls, but the order of the treatments and which texts will be assigned the treatment and which will be assigned the control will be randomized. These two aspects (order and selection of treatment) will act as control variables, in order to check whether they impact results.

B. Key data sources

The key data source for this experiment will be the answers by the participants in the experiment. Furthermore, the texts presented to the participants are taken from the *Parlspeech V2* dataset (Rauh and Schwalbach, 2020), which contains over thirty years of transcripts of speeches from the UK House of Commons, which are transcribed verbatim.

C. Hypotheses tested in causal chain

The key outcome of interest is the effect of simple language in politics on voting behavior. The main hypothesis asserts that when political parties use simple rather than complex language, this will increase support for this party. This is measured by the main survey questions posed directly after the participants read each texts, most specifically by the second question ("Would you vote for this person?").

Further outcomes of interest pertain to the effects of social class, education and party affiliation on the results.

The social class hypothesis asserts that the main effect will be stronger among people from a lower social class, as simple language can be considered an elite cue aimed at lower-class voters. This is measured through the social characteristics questions asked after the experiment's completion (2 questions, modelled after similar questions posed in the British Election Study).

The education hypothesis asserts that the main effect will be stronger among people who have experienced lower levels of education, as simple language can be considered an elite cue aimed at bridging the perceived gap between lower-educated voters and politicians. This is measured through the social characteristics questions asked after the experiment's completion (2 questions, modelled after similar questions posed in the British Election Study and following the guidelines by Advance HE, 2020).

The party affiliation hypothesis asserts that the main effect will be stronger among people who feel affiliated to the Labour Party, as this party has a traditional link with the lower-class, lower-educated voters who are regularly seen to display the highest levels of dissatisfaction with politics and feelings of a gap between them and politicians. This is measured through the social characteristics questions asked after the experiment's completion (2 questions, modelled after similar questions posed in

the British Election Study), as well as through one survey question, posed immediately after reading each text ("What political party do you think this person is affiliated with?").

D. Variable construction

The political distance variable (taken from the first survey question asked after each text, "Do you feel this person understands people like you?") will be a dummy variable, with the answer "Yes" coded as 1 and the answer "No" as 0. The answer "Don't know" will be coded as 99, similar to the coding procedure of the Eurobarometer.

The political support variable (taken from the second survey question asked after each text, "Would you vote for this person") will be a dummy variable, with the answer "Yes" coded as 1 and the answer "No" as 0. The answer "Don't know" will be coded as 99.

The speaker-party affiliation variable (taken from the third survey question asked after each text, "What political party do you think this person is affiliated with?"), will be a categorical variable, with "Conservative Party" coded as 1, "Labour Party" as 2, "Liberal Democrats" as 3, "Scottish National Party (SNP)" as 4, "Plaid Cymru" as 5, "Other" as 6, and "Don't know" as 99.

The post-experiment questions will deliver several additional variables, relating to the social characteristics of the respondent's self-identified social class, education levels and personal party affiliation. This will create categorical variables similar to the speaker-party affiliation variable described above.

For the first three variables described above (tied to the survey questions), respondents will be required to answer, which should take away the possibility of missing data on these variables.

For the post-experiment social characteristics questions, this requirement will not be in place. This could create missing data. Given the sample size of N = 200, the expectation is that enough data will still be supplied by the respondents. Imputation will not be used to address any of this potential missing data.

E. Treatment effect measurement

To measure the treatment effect, I will use a proportions test. I will take the proportion of individuals indicating that they would vote for the politician and compare it with the proportion of those indicating that they would not. I will do this for the entire group (all participants), to create a baseline value. After that I will compare the proportions of vote/not vote according to the treatments of simple versus complex language, and altered versus original text.
I will use a significance level of 0.5 as the basis for deciding whether or not to reject the null hypothesis.

F. How this project deals with multiple hypothesis-testing

This project contains one main hypothesis. The other hypotheses (social class, education, and party affiliation) are additional, but do not pertain to the main goal of the study. This means that there is no expectation of a multiple hypothesis-testing issue arising.

G. How attrition is addressed

As the study is estimated to take a maximum of 15 minutes for a participant to complete, the possibility of attrition should be limited. Moreover, the financial incentive is placed in order to ensure that, once started, a participant will complete the study. If a participant does not complete the study, the financial reward will not be paid out, something which is explained to participants prior to participation (through the informed consent form).
That said, attrition can always take place. This is why the sample size of N = 333 should leave enough room in place; even if a number of participants drop out of the study, the minimum required sample size is still met, as this is much lower (N = 64). Furthermore, the most likely point where participants may choose to stop participating, is during the social characteristics stage of the experiment, which takes place after the experiment. This means that attrition at this stage does not significantly impact the results on the main hypothesis, as this was already tested during the experiment.

H. How the study deals with outcomes of limited variation (small effect)

As variables with minimal variation might create noise in the analysis, these variables (with 95% of the observations displaying the same value) will be omitted from the analysis. If too many variables will have to be omitted because of this, the treatment effect in this study cannot be calculated, meaning the study does not yield any results.

Appendix F – Example experiment

Experiment OE_DPIR_PS

The purpose of the experiment is to investigate effects of political communication. While several studies have shown that politicians try to "speak" to an audience by adjusting their language to what they believe the audience wants to hear, this does not mean they are successful in doing so. They may not know what their audience wants to hear, or they may need to convince the audience of something unpopular. Furthermore, an audience may not be receptive to messages from certain politicians because of their own living situation or political ideology.

You will be asked to read a few speeches by politicians. All were delivered during a debate in Parliament. Some were slightly modified in order to increase readability, but the overall content of the speeches has not been changed. None of the speeches you will read are about topics currently being discussed in Parliament.

After reading each text, you will be asked the same set of questions:

1. Do you feel this person understands people like you?
   (YES, NO, DON'T KNOW)
2. Would you vote for this person?
   (YES, NO, DON'T KNOW)
3. What political party do you think this person is affiliated with?
   (CONSERVATIVE PARTY, LABOUR PARTY, LIBERAL DEMOCRATS, SCOTTISH NATIONAL PARTY (SNP), PLAID CYMRU, GREEN PARTY, UNITED KINGDOM INDEPENDENCE PARTY (UKIP), BREXIT PARTY, OTHER, NONE)

After the experiment, you will be asked to answer a few questions about yourself. You do not need to answer these if you feel uncomfortable doing so.

Please read and consider the text below carefully. You should spend at least one minute on this page. Only when a minute has passed will the submit button appear and will you be able to move on to the next page. When you have finished reviewing the text, please respond to the questions that appear on the remainder of the page in the way that best represents your views.

"No doubt the Secretary of State will confirm that, following the introduction of universal credit, when people's incomes change they will have to go to the local council to sort out their council tax benefit changes, and to the DWP to sort out their housing benefit changes. Two visits, or two contacts, will be required as a result of one change of income. What progress is the Department making in discussing with councils the need to provide a joined-up service so that, in future, people will need make only one contact when their incomes change?"

1. Do you feel this person understands people like you?
   - YES
   - NO
   - DON'T KNOW

2. Would you vote for this person?
   - YES
   - NO
   - DON'T KNOW

3. What political party do you think this person is affiliated with?
   - CONSERVATIVE PARTY
   - LABOUR PARTY
   - LIBERAL DEMOCRATS
   - SCOTTISH NATIONAL PARTY (SNP)
   - PLAID CYMRU
   - GREEN PARTY
   - UNITED KINGDOM INDEPENDENCE PARTY (UKIP)
   - BREXIT PARTY
   - OTHER
   - NONE

Please read and consider the text below carefully. You should spend at least one minute on this page. Only when a minute has passed will the submit button appear and will you be able to move on to the next page. When you have finished reviewing the text, please respond to the questions that appear on the remainder of the page in the way that best represents your views.

"An important problem with jobs for some people is that temporary jobs are available. Those are good jobs. They pay well. But they are temporary. They last only three or six months. Right now, if someone takes such a job, they would be unemployed again at the end of the temporary period. They would then lose the benefits that they had built up before. How will we make sure that people are not forced to take temporary jobs and end up in a worse position when the job ends?"

4. Do you feel this person understands people like you?
   - YES
   - NO
   - DON'T KNOW

5. Would you vote for this person?
   - YES
   - NO
   - DON'T KNOW

6. What political party do you think this person is affiliated with?
   - CONSERVATIVE PARTY
   - LABOUR PARTY
   - LIBERAL DEMOCRATS
   - SCOTTISH NATIONAL PARTY (SNP)
   - PLAID CYMRU
   - GREEN PARTY
   - UNITED KINGDOM INDEPENDENCE PARTY (UKIP)
   - BREXIT PARTY
   - OTHER
   - NONE

Please read and consider the text below carefully. You should spend at least one minute on this page. Only when a minute has passed will the submit button appear and will you be able to move on to the next page. When you have finished reviewing the text, please respond to the questions that appear on the remainder of the page in the way that best represents your views.

"How is this updated arrangement going to assist people like Mr. Wyatt, a constituent of mine, who was asked at the Department of Employment office how much he expected to receive if he was fortunate enough to find employment? When Mr. Wyatt answered a figure of £120 per week, my 58-year old constituent found that his income support had been put to an end. It was then resumed only a few weeks later, and the payments had been reduced by £6 per week. What is the right honourable Gentleman going to do about the manner in which people like my constituent are discriminated against?"

7. Do you feel this person understands people like you?
   - YES
   - NO
   - DON'T KNOW

8. Would you vote for this person?
   - YES
   - NO
   - DON'T KNOW

9. What political party do you think this person is affiliated with?
   - CONSERVATIVE PARTY
   - LABOUR PARTY
   - LIBERAL DEMOCRATS
   - SCOTTISH NATIONAL PARTY (SNP)
   - PLAID CYMRU
   - GREEN PARTY
   - UNITED KINGDOM INDEPENDENCE PARTY (UKIP)
   - BREXIT PARTY
   - OTHER
   - NONE

Please read and consider the text below carefully. You should spend at least one minute on this page. Only when a minute has passed will the submit button appear and will you be able to move on to the next page. When you have finished reviewing the text, please respond to the questions that appear on the remainder of the page in the way that best represents your views.

"The Work programme is not an underperforming programme: 60% of people are off benefits. We continue to modify and improve it, which is only right. We have set up a best practice committee so that people can get even better. There is no underperformance. We are proud of the record. I will tell the right hon. Gentleman one thing: those people who have got jobs, whom he dismisses so discourteously, are very proud of what we have done."

10. Do you feel this person understands people like you?
    - YES
    - NO
    - DON'T KNOW

11. Would you vote for this person?
    - YES
    - NO
    - DON'T KNOW

12. What political party do you think this person is affiliated with?
    - CONSERVATIVE PARTY
    - LABOUR PARTY
    - LIBERAL DEMOCRATS
    - SCOTTISH NATIONAL PARTY (SNP)
    - PLAID CYMRU
    - GREEN PARTY
    - UNITED KINGDOM INDEPENDENCE PARTY (UKIP)
    - BREXIT PARTY
    - OTHER
    - NONE

These questions are about you. You do not need to answer them if you do not feel comfortable doing so.

1. What is your highest completed level of education?
    - PRIMARY SCHOOL
    - SECONDARY SCHOOL
    - SOME UNIVERSITY
    - UNDERGRADUATE DEGREE
    - POSTGRADUATE DEGREE
    - DON'T KNOW
    - NOT APPLICABLE
    - PREFER NOT TO SAY

2. Do any of your parents have any higher education qualifications such as a degree, diploma or certification of higher education?
    - NO
    - YES
    - DON'T KNOW
    - PREFER NOT TO SAY

3. Do you ever think of yourself as belonging to any particular class?
    - NO
    - YES, LOWER CLASS
    - YES, WORKING CLASS
    - YES, LOWER MIDDLE CLASS
    - YES, MIDDLE CLASS
    - YES, UPPER MIDDLE CLASS
    - YES, UPPER CLASS
    - DON'T KNOW
    - PREFER NOT TO SAY

4. When you were growing up would you say your family was middle class or working class?
    - LOWER CLASS

- WORKING CLASS
- LOWER MIDDLE CLASS
- MIDDLE CLASS
- UPPER MIDDLE CLASS
- UPPER CLASS
- DON'T KNOW
- PREFER NOT TO SAY

5. During the last general elections, which party did you vote for?
   - PREFER NOT TO SAY
   - DON'T KNOW
   - LABOUR PARTY
   - CONSERVATIVE PARTY
   - LIBERAL DEMOCRATS
   - SCOTTISH NATIONAL PARTY (SNP)
   - PLAID CYMRU
   - GREEN PARTY
   - UNITED KINGDOM INDEPENDENCE PARTY (UKIP)
   - BREXIT PARTY
   - OTHER
   - NONE
   - DID NOT VOTE

6. Generally speaking, do you think of yourself as Labour, Conservative, Liberal Democrat, Scottish National, Plaid Cymru) or what?
   - PREFER NOT TO SAY
   - DON'T KNOW
   - NO/NONE
   - LABOUR
   - CONSERVATIVE
   - LIBERAL DEMOCRAT
   - SCOTTISH NATIONAL PARTY (SNP)
   - PLAID CYMRU
   - GREEN PARTY

- UNITED KINGDOM INDEPENDENCE PARTY (UKIP)
- BREXIT PARTY
- OTHER
- CENTRAL/SOMEWHERE IN THE MIDDLE
- A MIXTURE/SOMWHERE BETWEEN VARIOUS PARTIES
- INDEPENDENT
- ANY PARTY I AGREE WITH AT THE TIME