

Detecting Discrimination

PhD Dissertation 2019 © Malte Dahl
ISBN 978-87-7209-322-2
Printed by SL grafik, Frederiksberg, Denmark (slgrafik.dk)

Detecting Discrimination

*How Group-based Biases Shape Economic and
Political Interactions: Five Empirical Contributions*

Malte Dahl

mrd@ifs.ku.dk
+45 6014 3755

Department of Political Science
University of Copenhagen
Denmark
September 2019

This dissertation is presented for the degree of
Doctor of Philosophy

Preface

This is an article-based dissertation. It consists of a frame (chapters 1-4) and five research articles. The research articles are self-contained. They answer individual research questions and each of the five articles can be read on their own.

- A Dahl, Malte & Krog, Niels (2018). 'Experimental evidence of discrimination in the labour market: Intersections between ethnicity, gender, and socio-economic status' (Subsequently referred to as 'Intersections')
- B Dahl, Malte (2019). 'Alike but different: How cultural distinctiveness shapes immigrant-origin minorities' access to the labour market' (Subsequently referred to as 'Alike but different')
- C Dahl, Malte & Dinesen, Peter Thisted & Schioler, Mikkel (2019). 'Who is responsive? How electoral incentives and candidate selection shape ethnocentric responsiveness' (Subsequently referred to as 'Who is responsive?')
- D Dahl, Malte & Nyrup, Jacob (2019). 'Candidate choice in a high-information setting: Do ascriptive characteristics shape candidates' electoral prospects?' (Subsequently referred to as 'Candidate choice')
- E Dahl, Malte (2019). 'Social desirability bias in conjoint experiments: What is the optimal design when studying sensitive topics?' (Subsequently referred to as 'Social desirability')

Acknowledgements

This dissertation would not have been possible without the help and support of colleagues, friends and family. A long list of people have contributed and improved this dissertation immensely; any shortcomings that surely remain are my own.

First of all, I am extremely grateful to Peter Thisted Dinesen, who supervised this project. Peter is not only a brilliant researcher, he is also the most generous and supportive advisor that anyone could ask for. I greatly admire how he approaches research with enormous intellectual curiosity, analytical talent and hard work. If God is in the details, Peter's meticulous approach to all aspects of the research process serves as inspirational proof of this old idiom. Over the years, Peter has read and commented on countless drafts of the papers contained within this dissertation and his insightful ideas and feedback have improved the quality of my work substantially.

I am also thankful to my colleagues at the Department of Political Science, both former and present, for their inspiration, questions and criticisms. When I began my dissertation in 2016, I was fortunate to enter a stimulating and thriving community of aspiring scholars working on various aspects of political behaviour. Martin Vinæs Larsen and Jens Olav Dahlgaard deserve special mention. As teachers, and later as colleagues, both sparked my interest in experiments, helped inspire the dissertation and provided me aid and encouragement along the way. Other colleagues to whom I owe thanks include Jonas Hansen, Rasmus Tue Pedersen, Kasper Møller Hansen, Karina Kosiara-Pedersen, Frederik Hjorth, Mogens Jin Pedersen, Lene Holm Pedersen and Asmus Leth Olsen. Too many others to name made an impact on my time at the department and I acknowledge them all.

I owe a great debt of gratitude to my extremely talented former and current PhD colleagues for helpful comments, support and good times. Wiebke Marie Junk, Clara Vandeweerd, Anders Woller Nielsen, Livia Rohrbach, Lasse Aaskoven, Jonas Lind, Jens van der Ploeg, Anne Bach Nielsen, Yevgeniy Golovchenko, Christoffer Pfeiffer Cappelen and Ma Yi deserve special acknowledgment. A special thank you to Benjamin Egerod, my office mate over the course of the last year and a half. Benjamin is the model of what a scholar and good colleague is supposed to be. I will miss our daily talks about B-movies, the best records of the 70s and causal inference.

Further, I feel deeply indebted to my co-authors outside the department. I am grateful to Niels Krog, my old friend and the co-author of the first article included in this dissertation. His hard work, creativity and optimism played a big role in shaping the initial ideas for this project. Also, a big thanks to our man in Oxford, Jacob Nyrup. It has been truly inspiring to see someone complete as much high quality work within such different subfields as fast as Jacob. I also would like to thank Mikkel Schiøler for his excellent and hard work that made the third article possible.

Throughout my PhD program, I have received great comments on my work from a number of brilliant international scholars to whom I am grateful. Numerous academics helped frame my thinking as I pursued this research and I would like to thank in particular Alexander Coppock, Eline de Rooij, Florian Foss, Ryan Enos, David Broockman, Christian Grose, Michael Gaddis, Arnfinn Midtbøen and Alex Hughes. Also, a debt of gratitude to Jack Citrin for hosting my stay at UC Berkeley.

I also owe special thanks to The Danish Institute for Human Rights for their

generous financial support – and a special thank you to Line Vikkelsø Slot and Nanna Margrethe Krusaa. I am also grateful to Augustinus Fonden, Oticon and to Reinholdt W. Jorck og Hustrus Fond for the financial support that made my stay at UC Berkeley possible.

I have spent most days using various packages in the open source program R and I would like to thank all the people who volunteer their spare time to develop these packages. There are too many to thank here, but I am grateful in particular to Thomas Leeper, Alexander Coppock and Hadley Wickham.

I also would like to thank my former students. It has been fun and challenging to engage with our very talented students at the department. Kasper Arabi and Lotte Andersen also deserve special mention for their invaluable research assistance.

Finally, and most importantly, I would like to thank my family – my parents and my brother – for a lifetime of unwavering support. I owe my deepest gratitude to my beautiful wife, Rebecca. Writing a dissertation can be strenuous, but Rebecca has been a source of endless support and encouragement – and a constant reminder that work is not life. When days at the office were frustrating and long, it was all left behind once I headed home to spend time with Rebecca. A heartfelt thank you for your unconditional love and support.

Contents

1	Introduction	1
	Research question	4
	Some terminology	5
	Empirical settings and context	7
	Individual papers	9
2	Theory, Previous Work and outstanding questions	11
	A brief introduction to the two empirical tracks	12
	Theoretical underpinnings	14
	Theoretical propositions	21
	Some outstanding questions	22
3	Experimental approaches to measuring discrimination	29
	Enduring methodological challenges	30
	The use of experiments for studies of discrimination	32
	Ethical considerations	38
4	Core results, limitations and implications	43
	Core results	44
	Limitations	57
	Implications	59
	Future research	62
	Summary	65
	References	67

Research Articles

A	Experimental evidence of discrimination in the labour market	83
	Supplementary material	114
B	Alike but different	125
	Supplementary material	155
C	Who is responsive?	159
	Supplementary material	192
D	Candidate choice in a high-information setting	207
	Supplementary material	231
E	Social desirability bias in conjoint experiments	235
	Supplementary material	258

Chapter 1. Introduction

SOCIAL group categories can serve as powerful heuristics that shape individuals' attitudes, behaviours and sociopolitical interactions. A long-standing body of literature suggests that the tendency to categorize people based on predominant social categories is a basic process of human cognition (Allport, Clark, and Pettigrew 1954; Tajfel et al. 1971; Fiske 1998). While the use of category-based knowledge can generate efficient inferences about people's beliefs, traits or behavioural patterns, social categorization has potential downstream negative consequences. For example, by motivating various forms of inter-group conflict and group-based biases such as discrimination, understood here as treating individuals unequally for illegitimate reasons.

In many sociopolitical interactions, individuals' immutable group categories *should* be invisible. There are often strong legal or normative arguments emphasizing why societal actors ought to exercise impartiality. According to the laws of universalistic treatment and meritocratic principles, democratic governance is expected to disregard citizens' social categories such as class, ethnicity, gender or religious affiliation (Lippert-Rasmussen 2014; Lipsky 1980). In fact, it is a core virtue of modern bureaucracy that citizens are '*subject to formal equality of treatment*' (Weber 1978, p. 225). Moreover, in line with traditional assumptions about politicians, we would expect them to be equally responsive to their constituents, unconditional of whether they share descriptive characteristics (Cain, Ferejohn, and Fiorina 1987; Swain 1995).

Yet, a growing body of social science literature indicates that citizens' social categories *is* a factor in shaping how they are treated by actors or institutions tasked with upholding meritocratic principles and norms of equality. This ranges from discrimination in the labour market based on job

applicants' race (Bertrand and Mullainathan 2004) or religion (Adida, Laitin, and Valfort 2016) to the disparate treatment of citizens by street-level bureaucrats (Hemker and Rink 2017; Pedersen, Stritch, and Thuesen 2018) and in-group favouritism among legislators (Butler and Broockman 2011; Mendez and Grose 2018) to racial profiling in the criminal justice system (Knox, Lowe, and Mummolo 2019). This is an interesting starting point from a political science research perspective: When and how do social group categories have material and political consequences due to inter-group biases? What are the underlying motivations behind the observed patterns? What can institutions and individuals do to change such behaviour?

Studying these questions is of immense importance for our comprehension of core questions in political science for several reasons. First, the social group categories that social scientists focus on are politically relevant precisely because power, resources and disadvantages are often allocated along the lines of ethnicity, religion, gender and class, etc. (e.g. Gilens (1996), Carnes (2013), Adida, Laitin, and Valfort (2016), and Grose (2011)). Thus, groups and group categories are components in structuring who gets what, when and how – the very essence of politics (Lasswell 1950).

Secondly, as political scientists we care about political interactions – acts in which people come together to make decisions about common resources – but we also care about the attitudes and behaviours that precede such interactions (Enos 2017, p. 13). Studying individual-level behaviour in the labour market (Vernby and Dancygier 2019) or studying the interactions between citizens and public officials (Hemker and Rink 2017) or legislators (Butler 2014) sheds light on the processes by which resources and political influence are distributed.

Thirdly, discrimination can have significant downstream effects on political behaviour. Experiences of discrimination are associated with divergent political behaviour and have been documented negatively affecting citizens' trust in government and their perceptions of fairness (Oskooii 2018), decreasing political efficacy and participation (Schildkraut 2005; Wong et al. 2011) and affecting party identity (Kuo, Malhotra, and Mo 2017) as well as group consciousness (Sanchez 2008). There is evidence demonstrating that when immigrant-origin minority groups are disadvantaged in their access to economic resources and political influence, it creates a powerful cocktail that can provoke violent conflicts (Dancygier 2010; Dancygier and Laitin 2014). Others have argued that

discrimination against immigrant-origin minorities can result in a 'self-reinforcing discrimination equilibrium' – a situation in which natives and minorities act negatively towards one another in mutually reinforcing ways – ultimately preventing integration (Adida, Laitin, and Valfort 2016).

Fourthly, the large-scale inflow and permanent settlement of immigrants in most Western European countries actualizes questions of inter-group conflict and discrimination. As Putnam (2007) underlined in his now famous acceptance speech when he was awarded the Johan Skytte Prize: *'One of the most important challenges facing modern societies, and at the same time one of our most significant opportunities, is the increase in ethnic and social heterogeneity in virtually all advanced countries. The most certain prediction that we can make about almost any modern society is that it will be more diverse a generation from now than it is today.'* In contemporary European societies, no social categories are as salient as ethnicity, a key focus of this dissertation. Finally, apart from its academic relevance, the magnitude of discrimination against immigrant-origin minorities is an integrated part of political debates on integration, debates that are often grounded in divergent experiences and personal anecdotes. Without neglecting important perspectives from people's day-to-day encounters, causal evidence on behavioural responses is essential to inform public debates and public policy.

Much of the academic debate over the underlying motivations of discriminatory behaviour centres on the rationality of decision-making actors. Discrimination is commonly understood through an individual-level theoretical framework, positing that it is driven by either personal preferences (so-called taste-based discrimination (Becker 1957)) or else is based on the notion that group categories serve as information about difficult-to-observe characteristics and that actors engage in discrimination as a way to deal with uncertainty and optimize decision-making (so-called statistical discrimination (Arrow et al. 1973)). Although it remains inherently difficult to elucidate and disentangle these mechanisms in empirical contexts, these workhorse models provide a general framework for how to think about discrimination. These perspectives, however, do not consider the important insights into the microfoundations and dynamics of inter-group conflict and stereotyping illustrated in the rich social psychology literature. This literature provides, for example, the notion that the

mere process of categorizing individuals according to social categories can be sufficient to trigger discrimination (Macrae and Bodenhausen 2000), that we use simplified representations of heterogeneous groups as heuristics, which can distort our perceptions of others (Kahneman et al. 1982; Gaertner and Dovidio 2005) or that most forms of group conflict can be regarded as manifestations of a basic predisposition towards group-based social hierarchies (Pratto, Sidanius, et al. 1994).

Research question

My aim in engaging with this literature is to explore when and how social categories significantly impact political and economic interactions. I do so in five research articles that follow two empirical tracks. In the first track, I explore how social group categories shape citizens' encounters with public managers and private employers in the Danish labour market. In the second track, I study the effect of group-based biases on the representation of politically underrepresented groups. The overarching research question structuring this dissertation is:

How do group-based biases shape economic and political interactions between salient social groups?

How-questions can be ambiguous, and the research question is here understood in two ways. As a first-order concern, I seek to identify the causal effects of group categories on interactions. That is, everything else being equal, does information that cues a group category affect behaviour between social actors in such a way that some group members are systematically better or worse off? A secondary focus considers the mechanisms by which potential group-based biases work. I enquire as to why and under what conditions some group cues result in disparities, whereas others do not. The five articles in the dissertation address different research questions and make individual contributions. Thus, while each article can be discussed on its own, together they shed light on the overarching research question from multiple empirical cases and test different theoretical propositions.

To answer the research question, I bring together insights from different research traditions to expand on empirical, theoretical and methodological debates. I build on empirical insights from the

literature on discrimination in the fields of applied economics and political behaviour to address the micro-foundations of group-based inequalities by focusing on individual-level interactions. I apply the theoretical framework asserting that discrimination can be due to both personal preferences (taste) and strategic (statistical) behaviour (Arrow et al. 1973; Becker 1957; Butler 2014; Broockman and Soltas 2019) and draw upon theoretical insights from social psychology to augment the theoretical underpinnings of discrimination. Specifically, I incorporate insights from theories of social categorization, which offers a way to understand the cognitive processes by which people place others into social groups and how it shapes behaviour (Macrae and Bodenhausen 2000; Fiske 1998). Finally, I engage with a budding methodological literature on how to draw valid causal inferences and disentangle the effects of group identities on social outcomes from other factors using experimental research designs. Across the articles, I confront common methodological challenges in existing work and engage with recent discussions on measurement and potential violations of assumptions in experiments on discrimination (Gaddis 2017; Sen and Wasow 2016; Butler and Homola 2017).

Some terminology

When studying how economic and political interactions are shaped by actors' social group categories, we are led to ask: What categories lead to differential treatment and when? This suggests two more basic questions: What are social groups? Why are they important to human beings at all? To answer these questions in their entirety is beyond the scope of this dissertation, but some aspects are well understood in the literature, and in Chapter Two I return to these issues in greater detail. For now, I clarify a few of the dissertation's key concepts.

Social group categories play an essential role in this dissertation. By social group categories, I refer to psychological prominent categories that we apply when perceiving and describing others. Group categories such as ethnicity, religion or gender serve as cognitive tools that 'segment, classify, and order' the social environment (Tajfel 2010, p. 112). Such categories provide group members with an identification that is relational. That is, social categories acquire their meaning in contrast

to other categories, and social categorization is central to the definition of individuals as similar to or different from members of other groups (Fiske 1998).

The cognitive process of social categorization supports the formation of *stereotypes* understood here as ‘*mental representations of real differences between groups [...] allowing easier and more efficient processing of information*’ (Hilton and Von Hippel 1996). Stereotypes are localized around group features that are the most distinctive, provide the greatest differentiation between groups, and that show the least within-group variation (Hilton and Von Hippel 1996; Fiske 1998). Stereotypes are there for a reason: we use them as heuristics for making efficient decisions. A central finding of the social psychology literature on groups, however, is that categorizing people by their social categories accentuates perceived between-group differences, while perceived similarities within groups are exaggerated (Dovidio 2010). Concisely, while stereotypes allow for efficient assessment of others, they can be erroneous and distort perceptions of objective reality.

I label people who share a social group category as *in-group* members, while non-members are labelled as *out-group* members. All individuals belong to various groups, with some more politically salient and important in structuring behaviour than others. Thus, when an in-group/out-group dimension is salient, we can expect actors to apply stereotypical thinking to a greater extent. If behaviour or attitudes, on average, are affected solely by whether participants interact with in-group or out-group members, they engage in *group-based biases*. While in-group favouritism and out-group derogation are distinct concepts (Dovidio 2010), I group both components as group-based biases, inspired by Butler (2014).

I apply a broad understanding of *interactions*, by referring to direct or indirect contact between one or more members of two distinct groups. Interactions have an interpersonal connotation, but can also be non-personal or even institutional, with people from different groups coming together without prior or direct contact. When people vote, for example, they are interacting with other voters to determine a common outcome, although there is no direct contact (Enos 2017, p. 13).

Finally, I define *discrimination* as ‘*differential treatment on the basis of membership of a socially salient group*’ (Lippert-Rasmussen 2014). In other words, members of a group are treated differently (i.e. less favourably) than members of another group with otherwise identical characteristics in the

same empirical situation (Bertrand and Duflo 2017). One advantage of this definition is that it aligns with most anti-discrimination laws that often lists a number of groups which are protected by those laws (e.g., groups determined by gender, religion, race or sexuality) (Lippert-Rasmussen 2014). It follows from the definition, that individuals who are discriminated against do not necessarily have to be poorly treated, but treated worse than otherwise comparable individuals.

Discrimination is understood as the behavioral component of group-based biases whereas stereotyping is the most cognitive component and prejudice is the most affective component (Fiske 1998; Eagly and Chaiken 1998). However, discrimination does not presume any underlying cause. It is important to note, that the experiments in this dissertation measures discrimination on the average. The single act of favouring one candidate over another (for example a majority over minority applicant candidate in the labour market) cannot be defined as discrimination per se, since this choice could be a result of coincidental preference.

Empirical settings and context

I address the overarching research question by focusing on two distinct empirical contexts: the labour market and the local municipality councils in Denmark. There are a number of reasons why combining these empirical contexts is worthwhile, and why Denmark serve as an interesting context.

First, examining the research question in various empirical contexts allows me to better understand the common theoretical underpinnings of group-based biases. In Chapter Two, I elaborate on the theoretical basis and outline some propositions that generalises across empirical contexts. Second, although I speak to two existing subfields in the dissertation's two tracks, the research articles face some overlapping theoretical and methodological issues. Specifically, questions regarding accurate measurement and causal identification cross-cut research within the two contexts.

Third, the two contexts are relevant because they have policy implications. Inclusion of immigrant-origin minorities in local political bodies and into the labour market are decisive components for

the successful integration of immigrant-origin minorities (Dancygier 2010). As noted by Dancygier et al. (2015, p. 704): *'One of the chief reasons attributed to immigrants' discontent is the inequality they experience in the labour market and in the political arena'*. Some immigrant-origin minority groups in Denmark face significant economic and political inequalities. While there are unquestionably complex and multifaceted influences for these trends, it is pertinent to explore to what extent discrimination is an attributing factor.

Finally, addressing the research question in a Danish context is important for a number of reasons. As noted by Dovidio (2010, p. 16) there is a need to broaden the research horizon on groups to contexts outside of the US and to study interactions *'between members of different ethnic and religious groups coming together in differing circumstances with different norms, and against the backdrop of different legal and political systems.'* Studying the research question in Denmark allows me to discover how theories developed in other contexts can be generalized to very different settings and helps in establishing the scope of the findings (Gerring 2011). Thus, the articles in this dissertation contribute by exploring the extent to which between-group biases are confined to specific racial or ethnic out-group constellations or whether this is a more general phenomenon. Many factors may lead us to question whether the findings of discrimination replicates in Denmark: the strong meritocratic norms, the high degree of social cohesion, equality and the presence of a large public sector. Moreover, inter-group conflict and ethnic inequality are relatively new features of politics in Denmark, a country that has historically been characterized as highly homogeneous, which makes ethnicity a salient and interesting category for exploration.

Individual papers

This dissertation consists of five independent research articles, three of which are co-authored.

- A Dahl, Malte & Krog, Niels (2018). 'Experimental evidence of discrimination in the labour market: Intersections between ethnicity, gender, and socio-economic status' (Subsequently referred to as 'Intersections') Published in *European Sociological Review* 34 (4), 402-417
- B Dahl, Malte (2019). 'Alike but different: How cultural distinctiveness shapes immigrant-origin minorities' access to the labour market' (Subsequently referred to as 'Alike but different') Under review in *Ethnic and Racial Studies*
- C Dahl, Malte & Dinesen, Peter Thisted & Schioler, Mikkel (2019). 'Who is responsive? How electoral incentives and candidate selection shape ethnocentric responsiveness' (Subsequently referred to as 'Who is responsive?') *Paper presented at MPSA 2019*
- D Dahl, Malte & Nystrup, Jacob (2019). 'Candidate choice in a high-information setting: Do ascriptive characteristics shape candidates' electoral prospects?' (Subsequently referred to as 'Candidate choice') *Paper presented at DPSA 2018*
- E Dahl, Malte (2019). 'Social desirability bias in conjoint experiments: What is the optimal design when studying sensitive topics?' (Subsequently referred to as 'Social desirability') Under review in *Political Analysis*

The dissertation proceeds as follows: In Chapter 2, I lay out the overall theoretical framework of the dissertation and outline some outstanding questions. In Chapter 3, I consider a number of methodological challenges related to the study of how group categories shape interactions and subsequently discuss the advantages and assumptions of the research designs that I use in the articles. Moreover, I elaborate on the ethical considerations related to the studies. In Chapter 4, I discuss the contributions and limitations of the dissertation and outline relevant avenues for future research. Finally, prior to the inclusion of the five articles, I provide a brief summary.

Chapter 2: Theory, previous work and outstanding questions

IN this chapter, I outline the dissertation's theoretical framework and identify some important outstanding questions in the existing literature. I begin with a brief introduction to the relevant literature and consider the key empirical findings that have guided my research. I then present the theoretical foundation of why social categories can result in biased behaviour. Building on the two workhorse models of discrimination and drawing from social cognitive perspectives on inter-group behaviour, I present several theoretical insights that can deepen our understanding of the empirics. I summarize the theoretical underpinnings and observable implications in six propositions. Lastly, I identify the outstanding empirical, theoretical and methodological questions that I seek to address in the five research articles.

A brief introduction to the two empirical tracks

While this dissertation’s two empirical tracks derive from different academic fields, the five articles are tied together by similar theoretical underpinnings, as well as a focus on studying individual-level attitudes and behaviour with experimental research designs. This section situates the dissertation in the literature by briefly reviewing the key empirical findings that have influenced the five research articles.

In the dissertation’s first empirical track, articles A and B, I address whether and when labour market discrimination presents a barrier to immigrant-origin minorities’ hiring prospects. Questions of labour-market discrimination have traditionally been rooted in applied economics and sociology.¹ Specifically, much of the theoretical and methodological advancements stems from the field of labour economics (Aigner and Cain 1977; Altonji and Blank 1999; Guryan and Charles 2013; Bertrand, Chugh, and Mullainathan 2005; Becker 1957; Phelps 1972). A significant development in this literature was the turn away from model-based observational and cross-sectional work to experimental designs. In particular, the use of field experiments, like the *correspondence experiment* pushed the field forward (Bertrand and Dufo 2017; Guryan and Charles 2013). In a correspondence experiment in the labour market, the researcher sends a large number of job applications in response to job advertisements and randomly assigns a trait (e.g. race) of the job applicant. Potential discrimination is then identified by estimating the outcomes (invitations to interviews) for the fictitious applicants with and without the trait. The most common way to manipulate the perceived minority trait has been through the names of the applicants (e.g. female names, African-American names, Arabic Names, etc). Consistently – and with remarkably large treatment effects – this literature has demonstrated the acute breadth and magnitude of discrimination against racial, ethnic, and religious minorities (Vernby and Dancygier 2019; Adida, Laitin, and Valfort 2016; Bertrand and Mullainathan 2004; Midtbøen 2016). See Zschirnt and Ruedin (2016) and Quillian et al. (2017) for reviews of this literature.

¹ Political scientists have increasingly explored questions of labour market discrimination (e.g. Adida, Laitin, and Valfort (2016) and Vernby and Dancygier (2019)) and it has been a question of interest in public administration research (e.g. Villadsen and Wulff (2018)).

The use of experimental methods to uncover group-based biases has been widely applied by social scientists working on related questions in other empirical contexts (for a review, see Pager and Shepherd (2008) and Bertrand and Duflo (2017)). Particularly relevant for this dissertation, a growing body of political science research has applied experiments to examine biases in political institutions and its effects on representation (Grose 2014). In the dissertation's second empirical track, I engage with this strand of literature to explore when and how group-based biases shape political representation, and under which circumstances parity in political representation can be obtained.

Inspired by the correspondence study methodology, political scientists have explored legislators' provision of constituency communication and service (i.e. assistance that is not premised on voters' partisanship or typical political support). In the first field experimental study of race and representation in legislatures, Butler and Broockman (2011) uncovered how legislators exhibit racial biases in responsiveness to their constituents. Using a correspondence study design, the authors emailed requests for assistance to US state legislators and randomly assigned the constituents' race and partisan affiliation. Other studies have extended this work and consistently documented that politicians are more likely to advance the preferences of constituents with whom they share ethnic or racial characteristics (Butler 2014; Gell-Redman et al. 2018; McClendon 2016); that this behaviour is not exclusively explained by strategic considerations (Broockman 2013); and that biases in responsiveness is associated with policy decisions of elected officials (Mendez and Grose 2018). As such, constituency service experiments are powerful tools that enable researchers to uncover political inequalities. The findings from these experiments indicate that groups that are numerically underrepresented in office in the US are at a disadvantage, which carries implications for traditional debates on representation (e.g. Mansbridge (1999) and Fenno (1978)) and call into question the quality and inclusiveness of democracy.

That legislators display a willingness to discriminate against out-groups support the broader argument that descriptive representation of minorities in legislatures shape how well they are represented (Broockman 2013; Mansbridge 1999). In most Western European countries, however, immigrant-origin minorities are significantly underrepresented in political bodies (Bloemraad 2013).

While existing research points to a multitude of explanations for this persistent political underrepresentation, one possible contributing factor is that the electoral prospects of immigrant-origin political candidates is hampered on election day because voters prefer racial or ethnic in-group candidates, all else equal – so-called ‘electoral discrimination thesis’ (Portmann and Stojanović 2019; Broockman and Soltas 2019). Measuring electoral discrimination, however, is a challenging task (Thrasher et al. 2017; Fisher et al. 2015). One way to study voter preferences over political candidates is by mimicking real-world elections in survey experimental settings. The candidate choice conjoint experiment has proved especially promising as a tool to evaluate how voters react to attributes of political candidates (e.g., Teele, Kalla, and Rosenbluth (2018), Carnes and Lupu (2016), Sen (2017), Kirkland and Coppock (2017), and Ono and Yamada (2016)). This is the focus of the last part of the dissertation, in which I explore voter preferences in a nationally representative sample of voters (article D) and examine whether social desirability bias is a concern that undermines results from conjoint experiments (article E).

Theoretical underpinnings

If group-based biases influence economic and political behaviour as profoundly as evidence suggests, what do theories tell us about the underlying mechanisms? The most important disagreement in the literature concerns what drives discriminatory behaviour. More specifically, much of the theoretical debate centres on the rationality of decision-making actors. The classic distinction of group-based discrimination based on strategic or taste-based decisions illuminates this controversy and provides a basic framework for thinking about the underpinnings of discrimination.

Two theories of discrimination

Taste-based discrimination

The first perspective asserts that discrimination is *taste-based* (sometimes denoted preference-based) and has its intellectual roots in the work of Becker (1957). According to this perspective, actors prefer not to interact with a particular group of people because they hold a “taste for discrimination”, resulting in a willingness to pay a premium to avoid the psychic costs of such interaction. In the labour market, for example, employers may refuse to hire out-group members or only hire them at lower wages than other, identically-productive employees (Becker 1957). Thus, enough discriminatory employers in the labour market will result in a wage differential between otherwise identically-productive in-group and out-group employees (Guryan and Charles 2013; Becker 1957). In the same vein, if co-workers or customers hold a distaste towards members of an out-group, it will have similar consequences.

Scholars have widely applied this theory of taste-based discrimination to understand discrimination in other empirical contexts outside the labour market. Political scientists have employed the theory to explain politicians’ discriminatory behaviour towards out-group constituents (Butler and Broockman 2011; Butler 2014), citizens’ preferences for specific groups in naturalisation decisions (Hainmueller and Hangartner 2013) and voters’ preferences for specific political candidates (Broockman and Soltas 2019; Fisher et al. 2015).

It is important to note that preferences are exogenous. That is, there is no accounting for how people form unfavourable attitudes in the original formulation of the theory (Stigler and Becker 1977). Overlooking the determinants of a distaste or antipathy towards specific groups, the theory falls short of explaining the factors that might shape preferences over time or in different contexts such as personal experiences (Danckert, Dinesen, and Sønderskov 2017; Hjorth 2017), evolutionary perspectives (Aarøe, Petersen, and Arceneaux 2017) or a desire for group-based inequality (Pratto, Sidanius, et al. 1994). Nonetheless, the paramount implication is, that actors exhibit group-based biases not out of rational reasons, but because of their personal preferences.

Statistical discrimination

The alternative theoretical perspective presents a statistical (or information-based) model of discrimination. According to this perspective, discrimination is a signal-extraction problem that arises because actors with limited information turn to group categories to conjecture on difficult-to-observe characteristics (Phelps 1972; Arrow et al. 1973; Aigner and Cain 1977). There are two ways that group differences can result in statistical discrimination. First, two groups may differ on average. In the hiring process, for example, employers may assume that ethnic minority applicants compared to majority applicants are on average less qualified due to factors otherwise unindicated in job applications, such as limited language abilities or educational disparities. This makes it efficient to prefer the majority candidate, all else equal. Second, the groups may differ on group variances on some productivity-relevant variable. If minority applicants' productivity is characterized by greater variance, risk averse employers will prefer the majority candidate. In other words, actors rely on observable group categories to extrapolate the characteristics of individual group members, especially if it is costly to gain more information. Thus, discrimination serves as a way to cope with uncertainty and optimize decision-making (Arrow et al. 1973; Guryan and Charles 2013).

Though it originates in the context of the labour market, the theory of statistical discrimination is also widely applied to other empirical contexts. For example, in the literature on politicians' interactions with voters, it is commonly assumed that reelection-motivated politicians, for rational strategic reasons, invest less effort in constituent groups they presume are less likely to support them in elections (Christensen and Ejdemyr 2017; Fenno 1978). Politicians will focus on winning the most votes they can while expending the least amount of resources and because ethnic and racial groups engage in politics to various extents, biases against minority group voters is compatible with the notion of statistical discrimination. As such, the notion that legislators may exhibit biases against out-group constituents to maximize their personal vote is an efficiency argument parallel to how statistical discrimination is used in the labour market. In summary, according to the notion of statistical discrimination, actors might exhibit differential treatment because it is efficient.

Aligned with most scholars, I understand the statistical discrimination models to denote that, for discrimination to be classified as such, it must be grounded in accurate stereotypes (Aigner

and Cain 1977; Broockman and Soltas 2019; Akerlof 1976). In other words, evaluations of group differences should be based on actual differences and not "erroneous" perceptions (Aigner and Cain 1977)² – albeit there is some disagreement in the literature (e.g. Pager and Karafin (2009).) This understanding of statistical discrimination, however, does not address a central concern that stereotypes can be inaccurate (Bordalo et al. 2016; Bohren et al. 2019). A decision-maker may be worse off by relying on inferences about group-averages if they are grounded in inaccurate perceptions of the expected productivity or performance of a social group. One could argue that Becker (1957, p. 16-17) encompasses erroneous judgments in his model of taste-based discrimination: '*An employer may refuse to hire a [minority] solely because he erroneously underestimates their economic efficiency. (...) [A] taste for discrimination incorporates both prejudice and ignorance*'. However, if the discriminator cares about unobservables for reasons of efficiency, it is more consistent with the information-based account. This underlines a need for a conceptual distinction between *accurate* and *inaccurate* statistical discrimination and highlights the importance of including perspectives from psychology to augment our understanding of cognitive biases as a micro-foundation of discrimination.

Why distinguishing between the causes to discrimination is important and difficult

In order to strategically curb discriminatory practices, it is crucial to understand whether discrimination arises primarily from taste- or information-based dispositions. The solution to mitigate discrimination among employers, for example, will depend on which of these mechanisms best explains discriminatory behavior (Vernby and Dancygier 2019; Butler 2014). However, while conceptually distinct, the models of taste-based and statistical discrimination are not mutually exclusive and has proven difficult to convincingly distinguish in an empirical setting (Bertrand and Duflo 2017; Fryer and Levitt 2004; Butler and Broockman 2011). Internal motivations are difficult to measure empirically (Pager and Shepherd 2008) and, in most real-world settings, actors may engage in both taste-

² As Aigner and Cain (1977, p. 177) highlights '*a theory of discrimination based on employers' mistakes is even harder to accept than the explanation based on employers' "tastes for discrimination," because the "tastes" are at least presumed to provide a source of "psychic gain" to the discriminator. To interpret the "statistical theory of discrimination" as a theory of "erroneous" or "mistaken" behaviour by employers, as have some economists suggested, is therefore without foundation*'.

and statistical-based discrimination simultaneously (Hainmueller and Hangartner 2013). Moreover, even if one type of discrimination is successfully identified at one point in time, it can be a complex result of preceding processes. Disadvantages initially caused by taste-based discrimination can eventually initiate real group differences, creating a basis for statistical discrimination (Dancygier and Laitin 2014).

Despite these difficulties, many studies are able to provide suggestive evidence of the mechanisms underlying discrimination. This is often accomplished by manipulating pieces of information provided to individuals in real-world settings (Kaas and Manger 2012; Gell-Redman et al. 2018; Bertrand and Mullainathan 2004) or by eliminating strategic incentives in artificial settings (Hede-gaard and Tyran 2018). If discrimination is indeed a signal-extraction problem as presumed in the statistical discrimination theory, discrimination ought to be mitigated as positive information is provided to decision-makers. For example, field experiments on labour market discrimination have improved the credentials of the candidates to alleviate employers' reason to infer that minority applicants are deficient in some unobservable skills (Oreopoulos 2011; Bertrand and Mullainathan 2004; Nunley et al. 2016). Research on legislator discrimination has employed similar strategies, manipulating information like party-preferences (Butler and Broockman 2011) and intents to vote (Gell-Redman et al. 2018) to explore whether strategic considerations explain potential differential treatment. For example, Broockman (2013) cleverly demonstrates that black politicians are more likely to answer requests from black voters compared to their white counterparts, and that these differences increase when the sender claims to live outside of the politician's district (i.e., when there is no electoral gain to be had by answering). This behaviour can only be explained by nonstrategic motives (i.e. "tastes"; for example out-group prejudice, in-group favouritism or linked fate). This is an illustrative example of how a convincing identification of preference-based discrimination often relies on the premise that all strategic reasons for discriminatory behavior can be excluded.

Social cognitive perspectives on discrimination

On a parallel track, a burgeoning literature in social and political psychology have made considerable progress in understanding of group-based biases and their origins. A key insight is that the tendency to categorise people into social groups appears to be a basic process in human cognition. Part of a 'person perception process', this tendency prompts us to construct and employ categorical representations to make sense of other people and guide our behaviour (Macrae and Bodenhausen 2000; Fiske 1998; Turner 1982). Rather than considering individuals in terms of proclivities, we comprehend them on the basis of social categories.³

Social categorisation relates to how we process information in our two memory systems (Macrae and Bodenhausen 2000). On the one hand, humans have stable, general beliefs about the world that serve as standards against which we interpret experiences and stimuli – our “slow-learning” system. On the other hand, we can react to new experiences without automatically modifying our schematic knowledge. That is, we can respond to unexpected, novel stimuli – our “fast-learning” system (Baars et al. 1997; Kahneman et al. 1982). Social categorisation activates a range of long-term memory processes that then shape our expectations about others, including stereotypical judgments (Bodenhausen 1990). In other words, according to this dual process theory our most basic cognitive architecture deeply affect how we perceive and interact with other people (Mason 2018; Devine 1989). While social categorisation helps us navigate in a complex world, our general beliefs can sometimes prompt us to exceedingly rely on stereotypes. In the political behaviour literature, for example, ample evidence shows how heuristics guide voters' perception of political candidates (Lupia 1994; McDermott 1998).

From the social cognitive perspective we can derive three implications critical to this dissertation. First, because heuristics simplify representations of heterogeneous groups, this can lead to cognitive misconceptions and biases; that is, various systematic errors that in illogical and inaccu-

³ This social cognitive perspective stands in contrast to more instrumental notions that explain group-divisions as motivated by self-interest motives (Blumer 1958), for example over scarce resources (Scheve and Slaughter 2001) or because forming and maintaining favourable opinions of in-groups affirms self-esteem (Tajfel 1981; Turner 1982).

rate ways distort our perceptions of others (Kahneman et al. 1982; Gaertner and Dovidio 2005).⁴ For example, actors tend to assess a target group by overweighting its representative types (a so-called representativeness heuristic) (Kahneman et al. 1982; Bordalo et al. 2016). While stereotypes might contain a kernel of truth, we often exaggerate real differences and tend to evaluate out-groups with an unrealistic view of its relative merits (Tajfel 1981; Fiske and Neuberg 1990). Moreover, an individual's ability and motivation to process information affects these mechanisms. Factors like uncertainty, fatigue, or cognitive overload can further reinforce cognitive biases (Ma et al. 2013; Andersen and Guul 2018; Bodenhausen 1990). Thus, discrimination is not necessarily consistent with either the taste-based notion or the statistical discrimination models but can happen because people make spontaneous judgements with the aid of social categories and these can be inaccurate.

Second, the social cognitive perspective asserts that people categorise individuals based on multiple dimensions simultaneously and the strength of group-based biases depend on whether these dimensions are salient and whether they converge (Crisp, Hewstone, and Rubin 2001; Canan and Foroutan 2016; Crisp and Hewstone 2007). *Convergent bases* for categorisation occurs when an individual is perceived to belong to more than one salient category that indicates out-group membership. This increases inter-group biases (Canan and Foroutan 2016). *Divergent bases* for categorisation occurs when an individual is perceived to belong to conflicting categories, which indicates both (albeit conditional) in-group and out-group membership. This reduces negative stereotyping (Gaertner and Dovidio 2005). Thus, if an individual is perceived as simultaneously belonging to several out-group categories, out-group biases are more likely. This implies that some group categories and specific combinations of categories can motivate different behaviours. For example, Canan and Foroutan (2016) illustrate how Germans' perceptions of cultural dissimilarity with ethnic minorities are reinforced when introduced to a male, Muslim individual (as opposed to a female and/or non-Muslim individual). These findings align with a prominent line of political science scholarship that contends that out-group biases differ between out-groups, and that those biases are related to perceptions of cultural distinctiveness, visibility and salience of the out-group

⁴ In the same vein, unconscious psychological processes might impact behaviour. The notion of implicit biases asserts that unconscious and automatic psychological processes can result in biased behaviour (Bertrand, Chugh, and Mullainathan 2005; Devine 1989).

(Kinder and Sears 1981; Hainmueller and Hangartner 2013; Sniderman, Hagendoorn, and Prior 2004).⁵ Thus, on this basis, we can expect that group-based biases increase as individuals appear more socially distant and when multiple out-group categories coincide.

Third, stereotypes triggered by explicit social group categories may be affected by individuating information. Studies on the formation of impressions have documented how, if an individual is presented with information inconsistent with their held stereotype, individuating cognitive processes that counter stereotypes may be activated (Hastie and Kumar 1979; Pratto and Bargh 1991). Individuating information usually takes the form of traits or behaviors that contradict the stereotype of a specific social category (Chaiken and Trope 1999; Sherman, Macrae, and Bodenhausen 2000). Thus, we can expect that group-based biases can be mitigated by presenting individuating information that forcefully counters stereotypical beliefs.

Summary: Theoretical propositions

The previous sub-sections have outlined several perspectives on group-based biases and their driving factors. These perspectives, I argue, complement one another in the sense that several mechanisms might simultaneously contribute to group-based biases. The theories of taste-based and statistical discrimination serve as an overarching framework for how to think about discrimination. In brief, this framework contends that discrimination can be rooted in personal preferences or strategic behaviour. The social cognitive perspectives adds important perspectives. Specifically, because we tend to rely on heuristics rooted in simplified representations of heterogeneous groups, judgements of individuals are likely to be riddled by misconceptions and biases that distort perceptions of others. Furthermore, such biases are likely to vary in intensity and consequence depending on other characteristics of the individual, the perceiver, and the interaction between the two. The above perspectives can be summarized in six propositions:

1. Simple group-category cues can sufficiently and deeply affect human attitudes and behaviors,

⁵ This perspective contrasts the notion that negative attitudes towards immigrants reflect a generalized predisposition toward out-groups (Kinder and Kam 2010).

motivating group-based biases and prompting discriminatory practices.

2. Discrimination can result from rational decisions rooted in accurate information (statistical discrimination) and/or personal preferences (taste-based discrimination).
3. People spontaneously think with the aid of heuristics, which can cause error in judgement and promote discrimination rooted in inaccurate beliefs.
4. Group-based biases increase as individuals appear more socially distant and when multiple out-group categories coincide.
5. Individuating information that counters categorical representations can mitigate group-based biases.
6. Group-based biases have economic and political consequences.

It is important to note that the theoretical propositions have emerged through an inductive process in which the empirical findings also augmented the formulation of the theoretical expectations. As such, the six theoretical propositions derived from an iterative process. In the last part of the frame, I suggest directions for future research to test and advance these theoretical propositions.

Some outstanding questions

This section identifies some empirical, theoretical and methodological gaps in the existing literature to which this dissertation contributes. These questions guide as well as tie together the articles. Some of them cut across the two empirical contexts, whereas others address narrow debates within specific theoretical and methodological literature.

In the context of the labour market (Articles A and B):

Do immigrant-origin minorities face inequalities in their access to the Danish labour market?

As a first-order question, this dissertation aims to establish whether and to what extent immigrant-origin minorities face undue barriers in their access to the labour market. While research in other countries have addressed how racial or immigrant-origin minorities face discrimination (for a review, see Zschirnt and Ruedin (2016)), these questions have yet to be explored in the Danish context. The Danish welfare state lacks many of the institutional features that might otherwise trigger labour market discrimination (Epp, Maynard-Moody, and Haider-Markel 2017). Denmark’s formal hiring practices, large public sector, the high education levels of public sector employers and the egalitarian democracy suggest that discrimination might be a lesser concern in Denmark than in other countries. Articles A, *Intersections*, and B, *Alike but different* close this gap by addressing questions of discrimination in two field experiments involving a wide range of jobs in both the public and private sectors.

Different groups, different biases?

As a second-order question – and where a primary contribution to the literature more generally lies – the first two articles examine which and when out-group members become targets of discrimination. In articles A and B, I advocate the simultaneous study of several group categories because of the categories’ potentially mutually-reinforcing relationship. Using factorial designs, the two articles contribute to the existing literature by experimentally manipulating several group cues to elicit their effect alone and in combination.

As discussed earlier, one theory of what shapes perceptions of immigrant-origin minorities departs from the premise that majorities evaluate ethnic or racial out-groups based on their adherence to norms. The greater the perceived dissimilarity between immigrant-origin groups to the majority’s general cultural grounds, the more likely they are to be met with hostility (Schildkraut 2010; Wright and Citrin 2011; Hainmueller and Hopkins 2014; Ostfeld 2017). As Dancygier and Laitin (2014) observes, although scholars have developed increasingly sophisticated identification strategies to better understand attitudes towards out-groups, we know less about how and whether these sentiments matter for behavioural outcomes. Article B addresses this gap by randomly assigning group cues and information that taps perceptions of cultural distance.

Another outstanding question relates to the intersection between ethnicity *and* gender. The majority of correspondence studies on racial or ethnic labour market discrimination have refrained from manipulating both at the same time (the same is true in other empirical contexts). This perspective is incomplete because it assumes that discrimination is homogeneous across gender and existing scholarship may therefore present inaccurate accounts of ethnic disparities. While some research theorizes that immigrant-origin minority women are doubly and most disadvantaged (Harnois 2015; Ransford 1980; Beal 2008), from an evolutionary psychology perspective, researchers contend that ethnic conflict is most perpetuated by and most disadvantages men (Sidanius and Pratto 2001). Article A unfolds this broader theoretical disagreement on the intersection of gender and ethnicity.

Can individuating information reduce discrimination in the labour market?

Following from the theoretical propositions outlined previously, there is reason to expect discriminatory behaviour to be mitigated by individuating information (Fiske 1998; Pratto and Bargh 1991). Several studies on labour market discrimination have explored racial and ethnic discrimination across variations in the quality of candidates' credentials – such as educational background or past labour market experience – as a means to identify information-based discrimination (Bertrand and Mullainathan 2004; Oreopoulos 2011; Nunley et al. 2016). However, a cultural component of discrimination has been largely overlooked in the existing literature on discrimination (Rivera 2012). This raises a question of whether individuals who have adapted to their host countries' cultural norms and way of life obtain equality in access to the labour market as is sometimes claimed (Koopmans 2016). It also points to the question of what options immigrant-origin minorities have at their disposal to mitigate ethnic discrimination in the labour market. To test these conjectures, in article B, I theorize that discrimination can be mitigated by individuating information that refutes cultural-based stereotypes. I test this by examining the importance of both explicit and implicit information that tap into perceptions of cultural proximity.

In the context of local politics (Articles C, D and E):

Do local politicians discriminate against out-group voters? Can electoral incentives mitigate this behaviour?

Consistent evidence suggests that descriptive representation matters for how well politicians' represent their constituents (Iyer et al. 2012; Nye, Rainer, and Stratmann 2014; Chattopadhyay and Duflo 2004). One focus in the literature is how politicians provide constituency service and communication. The field experimental literature on political inequalities are based almost exclusively in the US (e.g., Butler (2014), Gell-Redman et al. (2018), Butler and Broockman (2011), Mendez and Grose (2018), and Janusz and Lajevardi (2016), but see McClendon (2016) for a notable exception). Thus, it remains an open question whether observed ethnocentrism in political responsiveness extends to contexts without racially-charged histories. Departing from the pervasive finding of ethnocentric responsiveness in the US, Article C, *Who is responsive*, thus considers whether the findings generalise to the Danish context. Moreover, speaking to the question of what motivates legislators to be more or less responsive to ethnic in- and out-groups, the article explores whether equality could be obtained if legislators face stronger electoral incentives. Previous studies have manipulated extrinsic components to behavior by randomly assigning information such as an intention to vote (Gell-Redman et al. 2018) or whether the requester is purported to live outside legislators' districts (Broockman 2013). We add to this literature by using both experimental and naturally occurring variations in electoral competitiveness to explore whether electoral incentives might curb discrimination.

Can voters select legislators who are responsive to their interests?

Departing from the pervasive finding of ethnocentric responsiveness it is pertinent to ask how, if at all, minority voters can improve the quality of political representation. Mendez and Grose (2018) suggest that legislators who favor public policies that adversely affect their minority constituents are more likely to be biased against ethnic minority constituents and less responsive to their requests. We build on this notion by exploring if minority constituents can identify politicians

more responsive to them by paying attention to parties and individual candidates' stated policy preferences vis-à-vis questions of immigration. To measure stated preferences, we utilize the fact that, prior to the experiment, a majority share of the incumbents seeking re-election participated in a publicly available voting advice application. By scraping this data, we retrieve information intended to inform voters about the positions of their local candidates on key political topics. The Danish multiparty system provides a good case for advancing the theory that parties' positions on issues concerning minority groups also shape interactions with minority voters. Because of the large number of parties, we can leverage greater variation in party positions related to questions of immigration and integration, from parties running on highly restrictionist immigration policies to parties that represent very liberal ones.

Do group-based biases shape the composition of political bodies?

Article D shifts focus from the behavior of politicians to the behavior of voters. Departing from the notion that underrepresentation of immigrant-origin minorities in legislatures affect their substantive representation, it is pertinent to explore reasons to the representational gap. While structural barriers to running for office clearly exist, it is less evident if immigrant-origin candidates face negative assessments from voters once they appear on the ballot. Research on this topic face methodological limitations because minority candidates run for office in different electoral districts, at different times, on different political platforms, with different party affiliations, and with varying campaign resources and experience (Broockman and Soltas 2019; Washington 2006). In article D, we implement a candidate choice conjoint experiment to elicit voter preferences across a hypothetical voting ballot. The conjoint experiment enables us to explore whether voters – both on average and across subgroups – exhibit preferences that benefit certain candidates directly or in more subtle ways. Survey experiments exploring the effects of candidates' ascriptive traits have typically done so in low-information settings where most information is held constant except the traits of hypothetical candidates (e.g. gender or race) (Banducci et al. 2008; Mo 2015; Philpot and Walton 2007). By providing voters with information that they often have access to in real elections, the design strengthens the external validity.

Can the candidate conjoint experiment be trusted when examining attitudes to sensitive topics?

One concern related to survey experimental studies – like the candidate choice conjoint experiment employed in article D – is respondents’ tendency to give untruthful answers because of social desirability. The ability to obtain reliable answers is a key inferential issue in the survey experimental literature, and it is well documented that respondents sometimes moderate their behavior in order to avoid the unease or distress that revealing socially undesirable answers may bring (Kaminska and Foulsham 2013; Kuklinski et al. 1997). Proponents of conjoint experiments often claim that these experiments sidestep issues of the social desirability bias (Hainmueller, Hopkins, and Yamamoto 2014; Horiuchi, Smith, and Yamamoto 2018; Teele, Kalla, and Rosenbluth 2018). However, despite a surge in political science research that use conjoint designs, little effort has been made to examine the extent to which social desirability plays a major role. In article E, I contribute to this literature by exploring the conditions under which social desirability is a concern by repeating two previous studies and testing the importance of design decisions.

Chapter 3: Experimental approaches to measuring discrimination

IN this chapter, I provide an overview of the data sources, outline characteristics of the dissertation's research designs and discuss some of their key assumptions. Given that the majority of this dissertation grapples with the question of how group categories shape economic and political interactions, I first outline empirical challenges to this query. Specifically, measuring the extent of group-based biases behaviour poses empirical challenges related to (i) establishing causality, (ii) obtaining accurate measures and (iii) external validity. I then discuss how the research articles in the dissertation address those challenges and under which assumptions. Specifically, I discuss the field experiments and the candidate choice conjoint experiments applied in the research articles. Finally, because the research articles in this project build on experiments that involved human subjects, I discuss the major ethical considerations and the ways in which those considerations affected the experiments.

Enduring methodological challenges

The research articles face three major methodological challenges. The first challenge relates to the establishment of *causality*: claims about how one phenomenon causes another. Early research on labour market discrimination identified discrimination by using individual-level outcome regressions that included observables for productivity and then interpreted the unexplained residual differential as a measure of discrimination (for a review of this literature, see Altonji and Blank (1999)).¹ It is, however, practically impossible to adequately account for all relevant control variables in this design (Heckman, Lyons, and Todd 2000; Guryan and Charles 2013). Further, since most variables of importance are often correlated with ascriptive characteristics such as gender or race, the control variables are measured post-treatment which is likely to introduce bias (Holland 1986). This calls for research designs that provide stronger causal identification, such as random assignment of some stimulus in randomized controlled experiments (e.g. Ayres and Siegelman (1995), Edelman, Luca, and Svirsky (2017)) or natural experiments (e.g. Hainmueller and Hangartner (2013), Tjaden, Schwemmer, and Khadjavi (2018), Enos (2017)).

A second important methodological challenge concerns the *accuracy of measures* retrieved from survey research. While studies on perceived discrimination are important in their own right, it is unclear to what extent perceptions of discrimination correspond to some reliable depiction of reality (Pager and Western 2012). Self-reported data on experiences of discrimination can be misinterpreted or overlooked leading to potential bias in estimates. Another well-known issue is that self-reported data from interviews or survey-based research may not elicit truthful responses to questions on sensitive topics – so-called *social desirability bias* (Wulff and Villadsen 2019; Tourangeau and Yan 2007; Hariri and Lassen 2017). Social desirability bias can be defined as ‘*a systematic error in self-report measures resulting from the desire of respondents to avoid embarrassment and project a favorable image to others.*’ (Tourangeau and Yan 2007). In other words, respondents tend to underreport socially undesirable activities and overreport socially desirable ones, resulting in distorted

¹ A particularly prominent method is the Oaxaca–Blinder decomposition, which separates differences (for example in average wages) into the part that is explained by differences in characteristics and the part that is explained by differences in returns to those characteristics and unexplained differences (Kline 2011).

measures. Social desirability bias is often considered as reflecting either *impression management* or a form of *self-deception* (Paulhus 1984). According to the impression management mechanism, survey respondents select the answer that is expected to maximize positive valuations by other subjects in their pursuit of social approval. The self-deception mechanism asserts that respondents provide untruthful answers in order to preserve and increase their sense of self-worth and minimize the cognitive dissonance resulting from the divergence between self-perception and their true preferences (Krumpal 2013). Although some findings indicate that anonymous and computer-mediated surveys reduce social desirability bias, evidence on social desirability bias in survey research generally indicates that it is a valid concern (Kuklinski et al. 1997; Janus 2010; Hariri and Lassen 2017) (see Gnambs and Kaspar (2017) for a review). Thus, while there have been important methodological innovations to address these inferential concerns (e.g. list experiments (Janus 2010) and conjoint experiments (Hainmueller, Hopkins, and Yamamoto 2014)), it remains a serious obstacle to the collection of accurate data.

A third, and to some extent related, methodological issue concerns *external validity*. Specifically, do causal estimates identified in a survey experimental setting accurately reflect decision-making, evaluation or behaviour in the real world? Because survey research is commonly carried out in artificial environments where responses have few apparent consequences, the results are less valid than studies that measure actual behaviour (Barabas and Jerit 2010; Gerber and Green 2012; Sears 1986).

Recent work by Wulff and Villadsen (2019), in which the authors externally validate two survey experiments against a field experiment, illustrates these inferential issues. In two seemingly realistic survey experiments, employers were asked by the authors to evaluate several job applications in which they had, without the participants' knowledge, randomly assigned the ethnic affiliation of the fictitious job candidates. Contrary to evidence from the field experimental study, employers generally preferred ethnic minority candidates in the survey experiments. This is supported by other research that indicates discrepancies between what employers say about their hiring decisions and their behaviour (Pager and Quillian 2005).

The use of experiments for studies of discrimination

The research articles in this dissertation explicitly grapples with the above-mentioned issues. The following sections outline how and under which assumptions these designs tackle the methodological challenges. To provide an overview, Table 1 contains a summary of the five research articles with respect to research design and data collection.

Table 1. Overview of the research design and data sources

Research article	Research design	Unit of analysis	Main data sources
A. <i>Intersections</i>	Field experiment	Individual	Experimental data
B. <i>Alike but different</i>	Field experiment	Individual	Experimental data
C. <i>Who is responsive?</i>	Field experiment	Individual	Experimental data; Voting advice application; Election data
D. <i>Candidate choice</i>	Conjoint experiments	Individual	Experimental data; Election data
E. <i>Social desirability</i>	Conjoint experiments	Individual	Experimental data

Conducting experiments in the field

In this section, I argue that the field experimental designs used in three of the research articles cleanly sidesteps the aforementioned methodological issues by leveraging a strong causal identification and high external validity. The random allocation in a well-designed experiment is one solution to the problem of unobserved confounders. By presenting respondents with carefully constructed and controlled comparisons, the experiment attains a high degree of internal validity (Gerber and Green 2012). Moreover, the field experiments alleviate a major concern of survey-based studies as well as survey and lab experiments more generally by measuring real world behaviour in a natural setting. The high external validity (as compared to for example a laboratory study) is a key advantage (Grose 2014; Teele 2014).

In articles A, B and C, I adopt a correspondence study design, a specific type of field experiment

in which the researcher audits real world behaviour among some subjects (e.g. employers or bureaucrats). Usually, this design involves the random assignment of some information (e.g. ascriptive traits of hypothetical candidates) to compare behaviour towards otherwise identical candidates.²

Core assumptions

The field experiments conducted as part of this dissertation rely on three core assumptions. The first assumption is that subjects are *randomly assigned* to either treatment or control with some known probability. In other words, if we define Y as our outcome of interest and d as the treatment variable, a subject, i , either receives treatment and reveals a treated potential outcome, $Y_i(d_i = 1)$, or receives control and reveals an untreated potential outcome, $Y_i(d_i = 0)$. Because we never simultaneously witness both the treated potential outcome *and* the untreated potential outcome, the treatment effect for an individual subject is an unobservable quantity. This also implies that we cannot infer the effect of the treatment for any individual subject (e.g. an individual employer). However, by virtue of their random assignment, the control and treatment groups are, in expectation, identical prior to that assignment (Gerber and Green 2012, p. 36). Therefore, we can identify the average treatment effect, ATE , across subjects by estimating the simple difference in means between treated and untreated subjects:

$$ATE = \overline{Y(d=1)} - \overline{Y(d=0)}$$

While it is impossible to randomly assign a person's ascriptive trait such as gender or ethnicity, it is possible to assign a trait that an experimental subject (e.g. an employer) *perceives* the job applicant to be. Note how this subtle difference in the research question – from ‘*What is the effect of a job applicant's ethnicity?*’ to ‘*What is the effect of the ethnicity an employer perceives a job applicant to have?*’ – allows for random assignment (Guryan and Charles 2013).

In the experiments employed in the research articles, randomization was always conducted using a random-number generator based on a seed to secure the reproducibility of the process (Coppock

² ‘Audit studies’ and ‘correspondent experiments’ are sometimes used interchangeably, while some define an audit study specifically as a study of real testers (auditors) matched for relevant personal characteristics.

2016). Articles B and C rely on a block-random assignment to achieve balance in the allocation of subjects to treatment arms across covariates (Gerber and Green 2012). For example, in article C we assigned incumbents to treatment groups using block randomization by political party, gender, the size of the municipality, incumbents' ethnicity and whether the incumbent was running for reelection. In article A, we tested for balance on observable covariates to test the robustness of our randomization scheme.

The second assumption is the *excludability assumption*. This assumption asserts that the potential outcomes are a function only of the treatment and not of some other feature of the assignment to treatment or by-products of the random assignment (Gerber and Green 2012). The excludability assumption was essential to the experiments in this dissertation. In fact, this assumption is crucial for all studies that rely on cues (e.g. names or pictures) as proxies for ascriptive group categories. More specifically, the excludability assumption asserts that potential differences in the subjects' response to a name is based exclusively on the signal that the name indicates about a particular group category (e.g. race or ethnicity). Thus, these studies implicitly assume that the experimental design isolates the effect of actors' 'racial' or 'ethnic' perceptions of a name's origin. Yet, because names have numerous connotations, this is not necessarily true. Therefore, discrimination attributed to distinctive names (e.g. ethnic minority names) might in fact be caused by a separate signal than ethnicity that the name also induces. The assumption is, however, only violated if subjects are affected by that unrelated information (Gerber and Green 2012). In other words, if subjects are only responding to the signal about ethnicity, then the excludability assumption still holds. I revisit this assumption in the articles, most explicitly in article A.

Furthermore, I bolster the results by using 'stimuli sampling' – that is, relying on many stimuli for a given manipulation to avoid potential idiosyncratic design choices that might affect the results (Gerber and Green 2012; Wells and Windschitl 1999). Instead of using just one name to indicate a group category, I always diversify the group proxy by using a large set of names. This also enables me to verify whether the pool of putative ethnic majority or minority names respectively yield the same treatment effects. Specifically, by regressing the outcome on the various aliases, in the articles

I demonstrate that there are no significant differences across minority names, which indicates that specific names are not mistakenly perceived as proxies for the majority group.

The third assumption is the *non-interference assumption*, also referred to as SUTVA (Athey and Imbens 2017). This assumption indicates that the potential outcomes for a subject, i , reflect only the treatment or control status of that subject without the treatment or control status for other subjects. Usually, the non-interference assumption is violated if treated and untreated subjects communicate with other treated and/or untreated subjects. For example, in article C, local incumbents received an email with highly identical questions. The non-interference assumption would be violated if legislators had discussed the requests with each other and accordingly changed behavior. Across the articles, the best way for me to check potential violations was to read the answers from all subjects, which did not indicate any reason to suspect a violation of the assumption.

The non-interference assumption is particularly relevant to article A, that employs a within-subject design. Instead of assigning each subject to *either* treatment or control, the individual employer received both (i.e. two applications that were comparable in quality and tone but varied in applicants' ascriptive characteristics). The advantage of the within-structure over between-subject designs is an increase in statistical precision. In this design, the non-interference assumption asserts that employers do not connect the two applications, an assumption that is rarely discussed but implicitly assumed in the literature.³ The risk of violating the assumption in correspondence studies is specifically a concern in designs that send multiple applications or unusual requests to subjects. Exactly for this reason, I rely on a between-subject design in research articles B and C. In article B, part of the treatment is the assignment of a CV photograph (of the same person) which would increase the risk of violating the assumption in a within-subject design. In study C, each legislator also received only one request. Had they received two or more comparable requests asking them similar questions, the design would have greatly increased the risk of interference.

³ Note that in article A the full schedule of potential outcomes remains unobserved because the two applications are not identical. We still observe either $Y_i(0)$ or $Y_i(1)$ for each subject (Gerber and Green 2012, p. 399).

In summary, field experiments are randomised, controlled experiments that takes place in the everyday environment of the subjects. This allows me to randomize and measure the effects of proxies that indicate some trait, e.g. ethnicity or gender, and convincingly estimate of potential discrimination. Under two additional assumptions – excludability and non-interference – the experiments provide unbiased estimates of the average treatment effects.

Candidate conjoint experiments for studies of group-based biases

The second research design that I use is the *candidate choice conjoint experiment*. Specifically, in articles D and E, I apply variations of this design. I constructed the experiments in Qualtrics software and they were distributed to a representative sample of Danish residents (article D) and through Amazon’s Mechanical Turk in the US (article E). While survey experiments face a number of limitations, some of which are outlined previously, they also offer a number of methodological strengths. Conjoint experiments specifically have been praised for their many advantages over ‘traditional’ survey experiments, placing them prominently within recent literature in political science. These experiments are effective and low-cost tools that enable researchers to explore respondents’ multi-dimensional preferences and test several causal hypotheses simultaneously (Hainmueller, Hopkins, and Yamamoto 2014; Hainmueller, Hangartner, and Yamamoto 2015). Hence, conjoint experiments have been leveraged in studies that explore attitudes towards immigrants (Hainmueller and Hopkins 2015), how voter preferences are shaped by political candidates’ gender (Teele, Kalla, and Rosenbluth 2018; Ono and Yamada 2016) or class (Carnes and Lupu 2016).

The candidate choice conjoint design juxtaposes pairs of hypothetical profiles featuring a combination of randomly assigned features that describe the candidates. This makes it possible to estimate the causal effect of multiple features simultaneously. The estimand of interest is typically the Average Marginal Component Effect (AMCE) which represents the average effect of a given feature level on the probability that the candidate will be chosen, averaged over the distribution of other features (Hainmueller, Hopkins, and Yamamoto 2014).

In research article D, we apply a candidate conjoint experiment to study voter preferences over candidates’ characteristics. This design addresses the concerns over the lack of external validity

and inaccurate measures in three ways. First, the candidate conjoint design relies on an analogy between the survey and a voting booth (Kirkland and Coppock 2017). Respondents choose between two hypothetical candidates based on randomly assigned information, which arguably reflects the process by which voters choose between political candidates. Thus, although responding to a survey is different from casting a ballot, an electoral choice may not be so different from a survey response. Secondly, an interesting feature of the AMCE is that it is defined as a function of the distribution of the treatment features. Therefore, we can explicitly control the target of the inference by including especially plausible or interesting features (i.e. information). For example, by defining the available information and its probability weights, it is possible to incorporate features that reflect a real world distribution (or other combinations of interest). This arguably increases external validity compared to classic survey experiments where only a few attributes are manipulated, while the broader political context is fixed (Hainmueller, Hopkins, and Yamamoto 2014).

Thirdly, a considerable advantage highlighted by proponents of conjoint experiments is that these designs have the potential to mitigate social desirability bias (Hainmueller, Hopkins, and Yamamoto 2014; Horiuchi, Smith, and Yamamoto 2018). As outlined previously, the ability to obtain reliable answers is a crucial inferential issue in survey experiments. The perceived ability of conjoint experimental designs to mitigate SDB is grounded in two notions. First, since respondents are presented with numerous features, a given sensitive feature is 'masked' among other features that are also randomly varied. Therefore, it is argued, respondents cannot infer that the sensitive feature is of particular importance (Teele, Kalla, and Rosenbluth 2018). Second, respondents can always find multiple justifications for any given choice (Hainmueller, Hangartner, and Yamamoto 2014). This implies that inappropriate answers can be justified by (combinations of) the levels of other features in the experiment. However, despite the prominence of conjoint designs, there has been surprisingly little empirical effort to examine the conditions under which social desirability bias is an issue. The main contribution of the research article E, *Social desirability bias*, is to qualify the extent to which candidate conjoint experiments provide accurate measures when respondents are asked to evaluate sensitive topics.

Ethical considerations

The field experiments conducted in three of the research articles have serious ethical implications that need to be properly addressed due to the involvement of human subjects and the element of deception. In this section, I consider the ethical aspects of the three field experiments (articles A, B and C) by focusing on the key ethical concerns related to the use of field experiments and the means that I took in order to minimize their ethical costs. Ethical considerations often address a question of gains (e.g. novel answers to important questions) and the extent to which they outweigh the costs (e.g. time costs imposed on subjects). This section focuses primarily on the cost-side of this equation, since the motivation for running the studies is detailed in other parts of the dissertation.

Ethical concerns in correspondence experiments

With the growing political science trend of the employment of field experiments as a research method (for an overview see Grose (2014) and Costa (2017)), concerns over ethics and research practices have received increasing attention (Desposato 2015; Teele 2014; Grose 2016). Especially important for this dissertation are the elements of *deception*, *consent* and *costs* imposed on subjects and the fact that the research involves public officials.

Deception is a common feature in many field experiments, including multiple studies on legislator-constituent relationships (Butler and Broockman 2011; Mendez and Grose 2018; Butler, Karpowitz, and Pope 2012) as well as studies on discrimination in the labour market (Midtbøen 2016; Adida, Laitin, and Valfort 2016). The element of deception is an integral part of the designs, albeit scholars have raised questions about deliberately deceiving subjects and discussed means to avoid this issue. For example, Butler, Karpowitz, and Pope (2012) implemented a field experiment in which they explored whether legislative offices are more responsive to service requests than they are to policy requests. To minimize the element of deception, the authors recruited confederates eligible to vote to write and sign the requests. Vries, Dinas, and Solaz (2015) build on the same idea in a field experiment on responsiveness among members of the EU Parliament in which they had voters send the requests to MEPs. These voters, who participated on a voluntary basis, were free to communicate further with their MEPs, but were asked not to disclose the fact that the initial email was part

of an experiment. The authors argue that *'[by] using real voters we do not have to invent aliases and mislead elected representatives'* (Vries, Dinas, and Solaz 2015). While the procedure of using actual constituent confederates to contact officials limits the element of deception, it may increase the burden faced by legislators through the demand for lengthy answers and further communication. In other words, there seems to be a trade-off between deception and time burden imposed on subjects.

Consent is another cost related to the use of field experiments. It is a general principle that research participants have the opportunity to provide informed consent based on sufficient information about the research, a principle that field experiments often violate. As explained previously, because ethnic or racial discrimination is socially undesirable behaviour, subjects may be unwilling to admit to discriminatory attitudes and thus behave differently if they are aware of their participation in this type of study (Wulff and Villadsen 2019). Thus, in experiments that examine behavioural responses to sensitive topics, the principle of consent is a major inferential concern (Riach and Rich 2004).

A third important issue to consider involves the assessment of the costs imposed on subjects. In terms of individual costs, most field experiments on discrimination have limited harmful effects on individual subjects. Answering a simple request can usually be done within a few minutes. However, if studies are conducted on large samples, the aggregated costs might be significant. For example, White, Nathan, and Faller (2015) contacted over 7,000 election officials in the US. Although each individual answer is short, the aggregated time spent on replies adds up to a significant amount. Moreover, the costs might be indirectly imposed on other citizens. Field experiments that involve public officials could take time that would otherwise be spent helping others.

Finally, specific considerations relate to the use of elite experiments that examine decision-making, responsiveness and representation in political institutions (Grose 2016). As McClendon (2016) points out in a thoughtful discussion of the ethics of experiments involving public officials, such research also introduces questions about researchers' ethical responsibilities to one another. The pool of politicians is small and recruiting elite subjects for experiments is a general challenge of the political science field (Druckman and Lupia 2012). Moreover, public officials often control

public research funding and research budgets. Thus, experiments comprised of elite subjects might jeopardise future research opportunities (Desposato 2015; Druckman, Leeper, and Mullinix 2014). Researchers should remain carefully attuned to this concern.

Addressing the ethical concerns in the research articles

Departing from the concerns outlined above, the field experiments employed in this dissertation represents a dilemma. On the one hand, the studies violate the principles of informed consent and involves an element of deception. On the other hand, this is an integral and necessary part in order to directly measure discriminatory behaviour. As Pager (2008, p. 78) argues, '*rigorous and realistic measurement of discrimination is fundamental to understanding and addressing persistent barriers facing members of stigmatized groups*'. In facing this dilemma, I had a clear obligation to minimize the burden imposed on the subjects and the ethical aspects of all studies were thoroughly considered.

In the field experiments conducted in the three articles, I took a number of steps to reduce costs to research participants. First, in the labour market studies (articles A and B), employers spent time reading job applications as part of the hiring process, and sometimes answered with a positive response or a rejection. It has therefore been a pivotal concern to reduce the time-cost as much as possible in all parts of the process. One way to do so was by handling all communication as fast as possible so that employers could quickly contact other potential candidates for the job. This was a priority, and I personally handled all communication to employers once a job application had been sent. This ensured that any callback was returned no more than a day later, and typically within a few hours, seven days of the week. In the study on political responsiveness (article C), politicians received a request sent to them by email. To minimize the time burden on individual legislators, we asked a simple, non-political question that they should have been able to answer immediately without spending resources looking for additional information. A median answer of 29 words reflects that the subjects were able to provide concise answers and the experiment required minimal time costs to the incumbents.

Second, another cost-reducing measure was ensuring anonymity and handling all data securely.

It was not my interest to name and shame individual subjects and by de-identifying subjects in the replication material, none of the experiments put any subjects at risk of exposure (Grose 2016). Across the projects, all data that contain the names of individuals and workplaces are kept in separate encrypted files on a protected server.

Third, before all three projects, a pilot study was conducted to ensure that the experimental set-up and the logistic of sending and handling the callbacks operated efficiently. Running the pilot studies was important and led to minor changes in our treatment (article C) and improved our infrastructure for handling answers (article A). Power calculations were conducted prior to the start of experiments. This played an important role in the designs in terms of deciding on the number of treatments as well as sample size. On the one hand, we wanted to ensure sufficient power in order to minimize the risk of not being able to identify potential population effects, which calls for large samples. Running an underpowered experiment clearly undermines its relative gains. On the other hand, there is no reason to expose more participants to the treatment than is necessary and thereby increase aggregated costs.

It is also worth pointing out that the experiments examined subjects in situations where they engaged in public rather than private behaviour (Grose 2016). As McClendon (2016) points out, communicating with voters is part of *'politicians' official duties'*. A key criterion for ethical evaluation on experiments that involve public officials is that researchers examine *'what is truly public behaviour by public officials or candidates, and not essentially private actions by public officials.'* (Grose 2016). A similar argument has been used to describe the hiring process in the labour market: hiring decisions are a subject of public interest and can be regarded as a public activity (Banton 1997; Zschirnt 2016). In the same vein, it is important to note that there is a legal dimension to the question of labour market discrimination, which is illegal under Danish law. In other words, the employers studied in articles A and B are obliged to follow principles of equal treatment.

The discussion of costs and gains that researchers face when considering how to conduct a field experiment is likely affected by two contextual aspects. First, the gains of field experiments are often *context-bound*. Consider the now canonical study by Bertrand and Mullainathan (2004) demonstrating the substantial discrimination faced by black Americans applying for jobs in the

American labour market. It is difficult to overestimate the importance of this study in informing the public, stimulating political debate on issues related to race and urging companies to reconsider their hiring practices. However, these gains are bounded to the US context with little spillover into policy debates in other countries. Further, the gains from fielding experiments on the same topic are arguably *marginally decreasing*. The first study that identifies discrimination is often more important than later replications, everything else being equal.⁴ Thus, one could argue that the fact that we had little knowledge of ethnic discrimination in the Danish context highlights the gains of running these studies. Finally, before collecting data, a request to have the ethical aspects of the project evaluated was sent to the Head of Studies at the Department of Political Science, University of Copenhagen. Based on a review of the literature and considerations of the project's ethical aspects, the project received acceptance to be carried out. Moreover, to ensure GDPR compliance, the project was approved by the responsible secretariat at the Faculty of Social Sciences.⁵

In summary, ethical considerations are crucial to this type of research. While there are clear and important ethical considerations, I argue that the field experiments are ethically defensible: they had very few harmful effects on individual subjects, the anonymity of individual subjects (and workplaces) was maintained, the experiments were characterized by methodological rigor and the studies focused on public subjects' official duties. Before all studies, I conducted power analysis to ensure sufficiently statistically powered samples. In other words, because the research was prepared and carried out carefully, there is only very limited detrimental effect on the subjects. At the same time, the question of discrimination is of high societal importance and the studies shed light on important normative issues. Moreover, there is no other way to credibly retrieve this vital information. Finally, I have argued that the novelty of the studies in the Danish context underscores their contribution to the field.

⁴ Replications are of course important: repeated studies over time can add to an overall assessment of whether discrimination is in decline. For example, a recent meta-study by Quillian et al. (2017) shows that discrimination is not falling.

⁵ The project was approved conditioned on some prerequisites: (i) The information is used only for the specified research purpose; (ii) that it is not possible to identify specific individuals; (iii) that all information is stored anonymously and appropriate technical and organizational measures are in place to protect personal data.

Chapter 4: Core results, limitations and implications

IN this chapter, I summarize the dissertation's contributions to the study of how group-based biases shape economic and political interactions. I begin by identifying the key insights to be drawn from the five research articles. This first part of the chapter is organized into two sections, corresponding to the two tracks of my empirical work. The first section outlines the main results from the research articles on how discrimination shapes access to the labour market, when, and for which groups. The second section presents the main results from the second track concerning how discrimination shapes political representation. Note that this chapter only summarizes the core results from the five research articles while more detailed results and analyses can be found in the individual research articles. I then discuss the limitations and broader implications of the main findings in the research articles. Finally, I point to avenues for future theoretical and empirical work on discrimination.

Core results

In the introduction of the dissertation, I formulated the following research question: *How do group-based biases shape economic and political interactions between salient social groups?* The next two sections briefly outline the main findings that speak to this question from each empirical track. Note that since the chapter draws on the research articles, there are overlaps between the articles and this chapter.

Group-based biases significantly shape access to the labour market

As discussed in the previous chapters, the literature on labour market discrimination generally indicates that discrimination against immigrant-origin minorities is persistent and widespread. Articles A, *Intersections*, and B, *Alike but different*, engage with this literature to identify the causal effects of group categories on hiring practices, and to consider how and under what conditions some group cues result in disparities, whereas others do not. The articles rely on field-experimental designs in which the ascriptive traits of fictitious job applicants were randomly assigned before sending the applications to a variety of employers advertising for a range of positions – a design that cleanly sidesteps issues of unmeasured confounding factors and allows for an assessment of real-world behaviour.

Collectively, the findings from the articles indicate that discrimination is a serious barrier and plays a role in shaping contemporary labor market disparities. Both research articles demonstrate that immigrant-origin minorities face labor-market barriers: job applicants with an immigrant-origin name receive significantly fewer invitations to job interviews compared to applicants with an ethnic majority name. This difference appears to be unrelated to discrimination based on socioeconomic characteristics. It is worth mentioning that the applicants had acquired educational merits, work experience and were written in flawless Danish. In addition to being the first studies to explore discrimination against immigrant-origin job applicants in the labour market in Denmark, they contribute to the literature more generally by advancing our understanding of the empirical patterns in several ways, which I outline in the following sections.

Discrimination is hierarchical

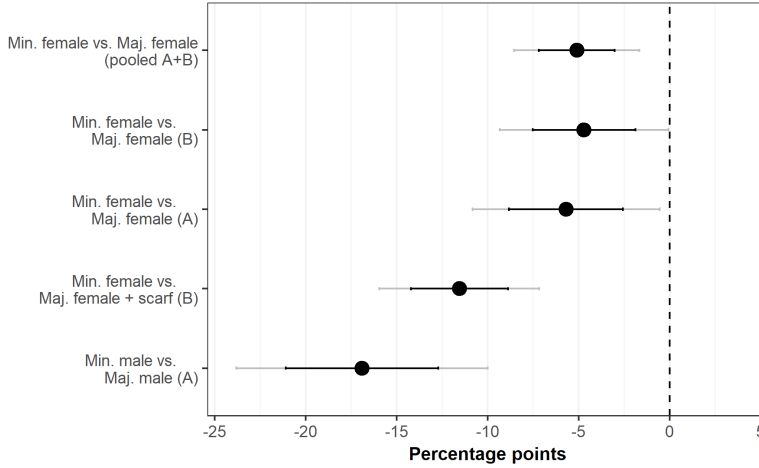
In Chapter Two, I proposed that we should expect group-based biases to increase as individuals appear more socially distant and when multiple out-group categories coincide. Articles A and B includes a variety of group configurations in order to test this proposition. That is, along the random assignment of ethnic affiliation, I manipulated gender, socioeconomic status, information about the applicants' spare time interests, and whether the applicants were depicted wearing a headscarf in a CV photo.

Figure 1 summarizes the results from articles A and B across the main treatment groups.¹ The figure depicts the average treatment effects (ATE) of having an immigrant-origin name ranging from 4.7 percentage points to 16.9 percentage points depending on the applicants' other characteristics. The figure reports five difference-in-means estimates using invitations to job interviews as the outcome variable. The first estimate depicts a difference between immigrant-origin minority and majority applicants of 5.1 percentage points for female applicants. This is a precision-weighted estimate derived from the average treatment effect of 5.7 percentage points in article A (second estimate) and an ATE of 4.7 percentage points in article B (third estimate). The fourth estimate (11.5 percentage points) indicates the ATE from comparing callbacks to immigrant-origin applicants wearing a headscarf with majority female candidates. The fifth and final estimate reveals a difference of 16.9 percentage points between male immigrant-origin minority applicants and male majority applicants.

Overall, discrimination does not appear to be homogeneous across out-groups, but rather depends on the specific configuration of group categories. Male immigrant-origin minorities are subject to significantly more discrimination than female immigrant-origin minorities, albeit when immigrant-origin females wear a headscarf, the effect of having a minority name is amplified. The results are consistent with the notion that stereotypes about ethnic minorities are mainly ascribed to males (Eagly and Kite 1987), findings from survey experiments that indicate that perceptions

¹ While one should generally be cautious when comparing results from correspondence experiments (results are sensitive to context, design, job types etc.), the experimental procedures in the two studies are almost identical; this is why they can be integrated into a combined analysis.

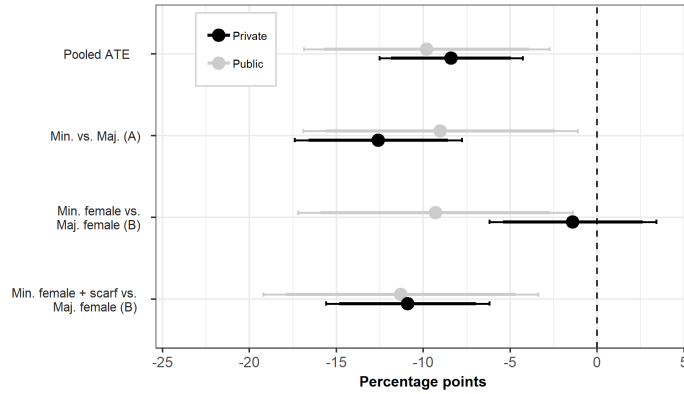
of cultural dissimilarity have negative effects on attitudes (Hainmueller and Hopkins 2015; Ostfeld 2017) and that perceptions of cultural dissimilarity with ethnic minorities are reinforced when subjects are introduced to a male and/or Muslim individual (as opposed to a female and/or non-Muslim individual) (Canan and Foroutan 2016; Adida, Laitin, and Valfort 2016). Notably, among majority applicants there is no differential treatment across gender which indicates that the gendered aspect of discrimination among immigrant minorities is not driven by a general preference for hiring women. On a more general level, the results are consistent with the notion that some group categories and specific combinations of categories can motivate different behaviours. The findings also underline an important methodological point, namely the importance of carefully considering proxy markers for minority group membership. Researchers examining discrimination against out-groups should attend to the fact that some characteristics strengthen the risk of discrimination.

Figure 1. Differences in callbacks between groups (articles (A) and (B))

Note: The Figure plots the average effect of having an immigrant-origin minority name as compared to a majority name on invitations to job interviews when applicants are female (estimates 1-3), wears a headscarf (estimate 4) and are males (estimate 5). The first estimate is a precision-weighted pooled estimate derived from articles A (2018) and B (2019). 90% and 95% CI are included. $N = 2150$.

Discrimination is consistent across sectors

During the process of collecting data, I read more than a thousand job ads. Frequently included in these ads was the statement that 'all applicants, unconditioned on their religious affiliation, ethnicity, gender or sexual orientation, are encouraged to apply.' In fact, this proposal is integrated in all public sector job ads. Nonetheless, as shown in Figure 2, discrimination is consistent across private and public sector jobs. A precision-weighted estimate that combines the results from articles A and B indicates that public and private employers exhibit biases against immigrant-origin minority applicants to roughly the same extent. This finding contradicts a belief, commonly held, that public sector organizations are more egalitarian and discriminate less than do private workplaces (see for example Midtbøen (2015)). It also raises questions regarding the effectiveness of organizational procedures such as formalized hiring procedures and the use of job templates – typical features of public sector hiring practices – as mechanisms for combatting discrimination in hiring as is often suggested in the literature (Pager and Shepherd 2008; Moulton 1990).

Figure 2. Average treatment effects across sectors in articles (A) and (B)

Note: The Figure plots the average treatment effect of having an immigrant-origin minority name as compared to a majority name on invitations to job interviews across public and private sector jobs. The first estimate is a precision-weighted pooled estimate derived from all treatment groups in articles A (2018) and B (2019). 90% and 95% CI are included. $N = 2395$.

Immutability of ethnic discrimination

Departing from the findings of discrimination, I explore what options, if any, immigrant-origin minorities have to avoid discrimination. In Chapter Two, I suggested that if individuating cognitive processes that counter stereotypes are activated, discrimination should be curtailed. To explore this notion, in research article B, I randomly assigned information about applicants' adherence to cultural norms (i.e. civil engagement, membership of local sports club, a relationship with a native Dane) and whether the immigrant-origin minority applicant wore a headscarf. Thus, the research article raises the question of whether discrimination is moderated by the extent to which a minority individual appears culturally deviant. The results in article B indicate that immigrant-origin minorities cannot affect their employment chances by providing implicit information that signals cultural adherence. While an explicit and visual signal (wearing a headscarf) increases discrimination significantly, applicants gain nothing by signalling that they are culturally proximate. A potential explanation for this null-finding is, that employers rely on a 'lexicographic search' whereby they stop reading once they see the applicants' name or picture and therefore fail to see all credentials (Bertrand and Mullainathan 2004). Nonetheless, this finding demonstrates that the

effects of fixed traits like ethnicity and gender are difficult to circumvent, and that immigrant-origin applicants have few tools at their disposal to escape discrimination.

Group-based biases and political representation

Articles C and D contribute to longstanding debates in political science regarding the factors shaping and the implications of descriptive representation. A basic principle of democracy is the claim that citizens' preferences should count equally in the realm of politics. However, as outlined in the following sections the research articles indicate that group-based biases significantly hamper immigrant-origin minorities access to political representation and influence in the context of Danish local municipality councils.

Politicians exhibit ethnocentric behaviour – even when they face strong electoral incentives

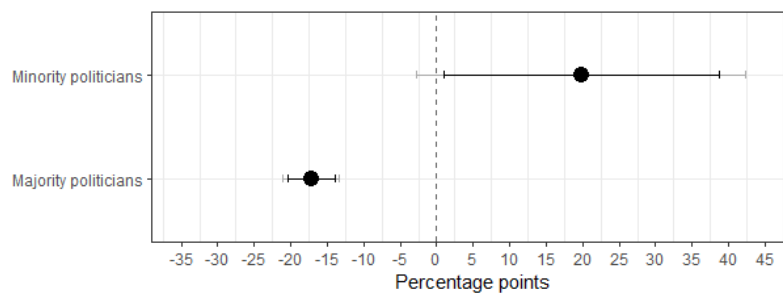
As discussed in Chapter Two, a U.S.-based body of literature indicates that politicians are more likely to answer constituents who share their ascriptive traits. Article C, *Who is responsive*, engages with this literature by exploring how Danish local incumbents respond to constituent requests asking for help with locating their polling station for the upcoming election. As depicted in Figure 3, immigrant-origin constituents are significantly less likely (17.2 percentage points) to receive an answer from ethnic majority incumbents, whereas the reverse pattern is true for immigrant-origin incumbents responding to ethnic majority constituents (19.8 percentage points). This also manifests in responses of lower quality to requests from out-group constituents.² Overall, this finding aligns with evidence from the American context (Mendez and Grose 2018; Butler 2014; Gell-Redman et al. 2018; Butler and Broockman 2011). The strong pattern of ethnocentrism in responsiveness among legislators in Danish local councils suggests that ethnocentric responsiveness is a more general phenomenon that extends to settings with much less ethnicized politics.

As previously discussed, recent work on representation points to two mechanisms that could explain why representatives are less responsive to immigrant-origin constituents (i.e. personal pref-

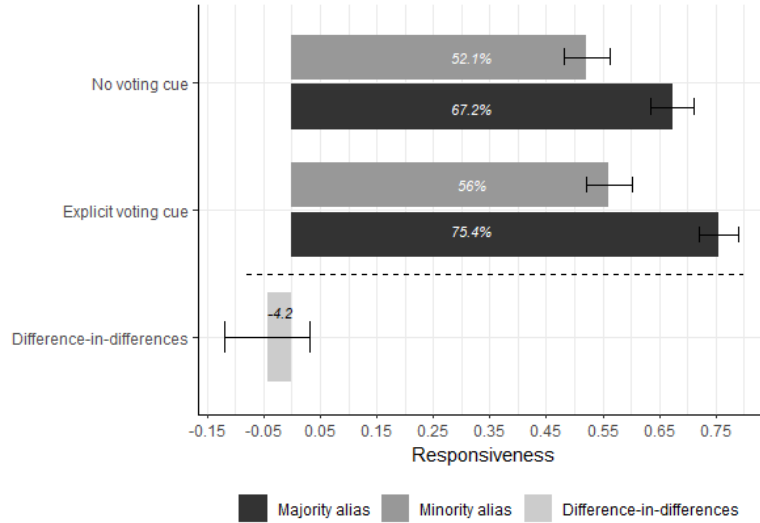
² Specifically, when the requester holds an immigrant-origin minority name, the replies are shorter, less timely, significantly less friendly, less likely to provide an answer to the question, and less likely to include an invitation to follow up.

erences and strategic motivations). Departing from this literature, we ask how, if at all, such tendencies can be curbed. More specifically, we scrutinize two different mechanisms for obtaining equal responsiveness among legislators. First, we test how majority legislators' responsiveness varied with increasingly strong electoral incentives by randomly assigning legislators an explicit expression of intent to vote in that legislator's favour. Moreover, we merge individual legislator responses with auxiliary data on their electoral performance (seat-winning margin) and their decision to run for re-election in an upcoming race. Overall, the results demonstrate that in-group favouritism in constituency service occurs even when there are clear strategic incentives for incumbents to respond to voters. Figure 4 shows that when an explicit statement of intention to vote for the legislator is included in the request, the likelihood of responsiveness increases; however, the differential treatment is not mitigated. If anything, the in-group favouritism increases slightly when given an explicit vote cue as indicated by a negative but statistically insignificant interaction term. The article demonstrates that this is true for other types of electoral incentives as well. This finding speaks directly against the notion, that if elections are sufficiently competitive it will animate office-seeking legislators to be responsive to their constituents independent of their ethnic background.

Figure 3. Effects of the immigrant-origin alias on responsiveness among ethnic majority and minority incumbents



Note: The figure plots the average treatment effect of signing a request to incumbents with an immigrant-origin name compared to an ethnic majority name. The outcome is receiving an answer. Treatment effects are provided for ethnic majority and ethnic minority politicians. $N = 2395$.

Figure 4. Responsiveness across constituents' alias and inclusion of explicit voting cue

Note: The figure plots the average response rate to voters with a majority or a minority alias conditional on whether a personal voting cue was included in the requests. Finally, the difference-in-differences between these groups is reported as percentage points. $N = 2395$.

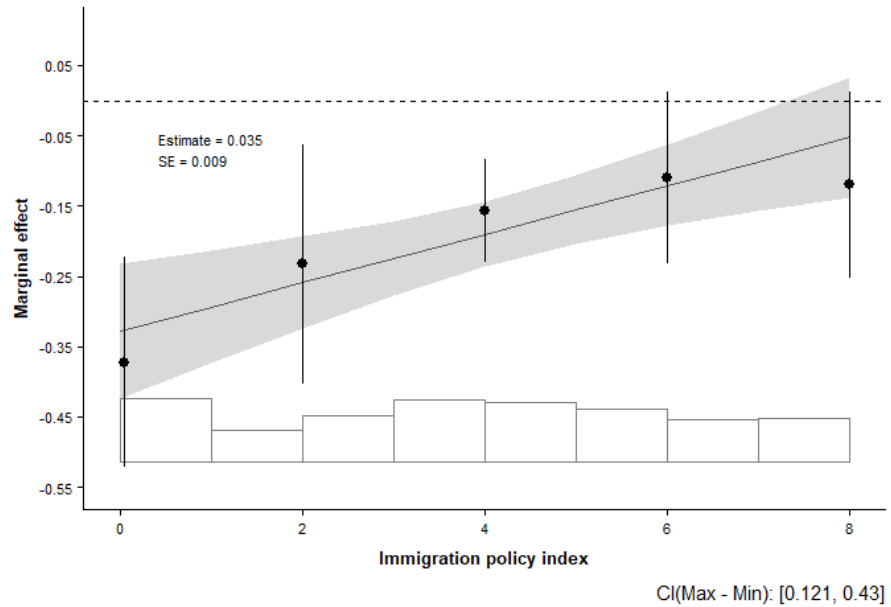
Incumbents' political preferences vis-à-vis immigration policies predict their personal biases

The third contribution of article C is an examination of the strategies by which minority voters can identify politicians who are more responsive to them. Specifically, we consider politicians' political preferences on questions that relate to integration and immigration policies by merging the experimental data with politicians' stated preferences, gauged via a voting advice application. Hence, we were able to examine whether politicians that explicitly favour a liberal immigration policy are also less likely to exhibit biases against immigrant-origin constituents.

Our findings provide compelling evidence that immigrant-origin constituents are able to identify politicians that are less discriminatory by paying attention to their policy preferences. We first demonstrate that immigrant-origin minority constituents are significantly more likely to be provided an answer when they contact incumbents from parties that advocate liberal immigration and integration policies. We then show that even among candidates within the same party, indi-

vidual legislators' stated preferences on immigration policies can be effectively used to locate more responsive politicians. Figure 5 plots the response-outcome regressed on an interaction between the immigrant-origin name treatment and the immigration policy measure with party fixed effects. The positive and significant interaction indicates that, even when differences in immigration-policy preferences between parties are accounted for, incumbents who explicitly favour a liberal immigration policy are less likely to exhibit a bias in responsiveness.

Figure 5. Marginal effect of immigrant-origin alias on politicians' responsiveness across the immigration policy measure (+ party FE)



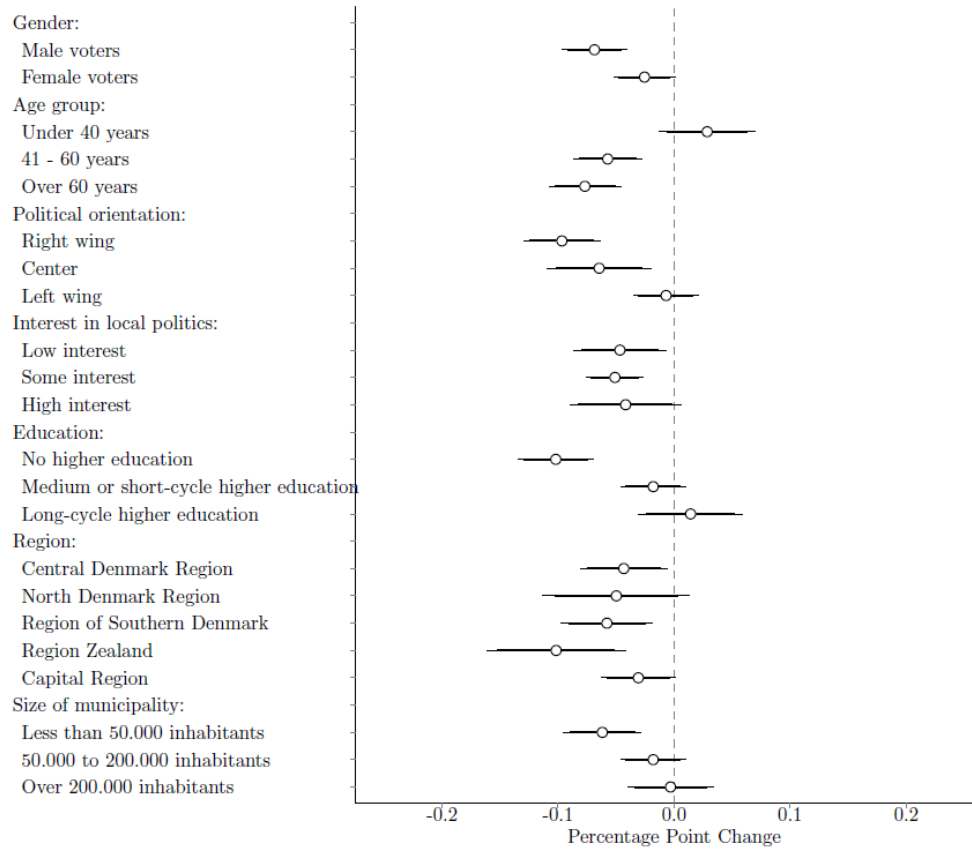
Note: The plot indicates the marginal effect of the minority alias across immigration policy, including party fixed effects. A score of 0 on the measure indicates a restrictive immigration profile. Bins with 95% CI are included. $N = 1514$.

Voter preferences shape the composition of political bodies

Article D, *Candidate choice*, applies a candidate choice conjoint experiment to explore if voters exhibit ethnic biases in evaluating political candidates for local municipality elections. While there is ample evidence to indicate that voters use candidates' ascriptive traits as information shortcuts to make inferences about candidates' policy preferences or personality traits (Kirkland and Coppock 2017; McDermott 1998), it is less certain if ascriptive traits matter when voters have access to ample information about the candidates as is often the case in real elections. To explore this notion, we invited a nationally representative sample of voters to participate in multiple 'elections' in the form of a candidate choice conjoint experiment. In this survey experimental setting, we mimicked real-world elections by asking voters to evaluate multiple pairs of hypothetical local political candidates described by several randomly assigned attributes.

The results indicate that, all else equal, immigrant-origin candidates are evaluated significantly worse compared to majority candidates. On average, voters favour the ethnic majority candidate by 5 percentage points. As depicted in Figure 6, this treatment effect is remarkably consistent across various subsets of the sample. Ethnic majority candidates are preferred across voter characteristics such as gender, political interest, and to some extent across party preferences and political leanings. This indicates that political candidates with immigrant-origin backgrounds face barriers because voters perceive them more negatively compared to ethnic majority candidates – even when voters are provided with information on the candidates political party, their preferences on salient policy questions, educational background, age and political experience.

Figure 6. The effect of an immigrant-origin alias compared to an ethnic majority alias on the likelihood of being the preferred candidate across voter characteristics



Note: The figure indicates the effect of having an immigrant-origin name compared to an ethnic majority name (baseline) on the likelihood of being preferred as candidate among voters. If the dot is to the left of the baseline, a majority candidate is preferred. $N = 1551$.

Can we trust the results from candidate conjoint experiments?

In addition to the dissertation's substantive contributions, article E offers an important methodological contribution. Departing from the findings in article D is the question of whether the results in conjoint experiments are biased due to social desirability. As discussed in Chapter Three, reliable answers is a key inferential issue in survey-based work and we might fear that respondents moder-

ated their behaviour in order to avoid the unease that revealing socially undesirable answers can bring. In article E, I tested the conditions under which social desirability is a concern in candidate choice conjoint experiments. Specifically, I ran two studies inspired by Sen (2017) and Hainmueller and Hopkins (2015), respectively. In each study, respondents were randomly assigned to one of three conjoint designs intended to either minimise or amplify their attention to sensitive dimensions and their possibilities for justifying inappropriate answers. Specifically, in the first condition, *the high-contrast paired design*, each respondent was presented with a number of conjoint pairs in which the levels of a sensitive feature were repeatedly contrasted (e.g., a black vs. a white candidate). The second condition, *the restricted paired design*, was similar, except that the sensitive feature was only contrasted in a limited number of conjoint pairs. Finally, the third condition was a fully randomised *single-profile design* showing only one candidate at a time.

The results demonstrate that these design differences significantly affect respondents' inferences about the research objective (i.e. their attention to a sensitive feature). As expected, respondents assigned to a high-contrast paired conjoint design are much more likely to infer that the sensitive feature is the main focus of the study than respondents assigned to either of the other two designs. Surprisingly, and most importantly, the design differences do not translate into any immediate effect on respondents' priorities. When comparing the effects of the sensitive features across designs, there are no distinguishable differences: respondents' answers are stable. This is true even among groups of respondents who are known to be more likely to provide socially desirable answers. Figure 7 and Figure 8 depicts each coefficient estimate for the high-contrast design versus estimates obtained from the restricted designs (left) and the single-profile designs (right). Each point represents an AMCE-estimate with 95 percent confidence intervals with the coefficients ordered by their magnitude from most negative to most positive. The sensitive feature-estimates are coloured blue. Thus, the figure visualizes the extent to which larger AMCEs in the high-contrast designs are associated with larger effects in the alternative designs. Altogether, there are no apparent differences in the AMCEs between the experimental conditions. This evidence indicates that when researchers use conjoint designs to study sensitive topics, they should not compromise their choice of design due to the fear of SDB.

Figure 7. Study inspired by Sen (2017). High-contrast estimates versus estimates from the alternative designs

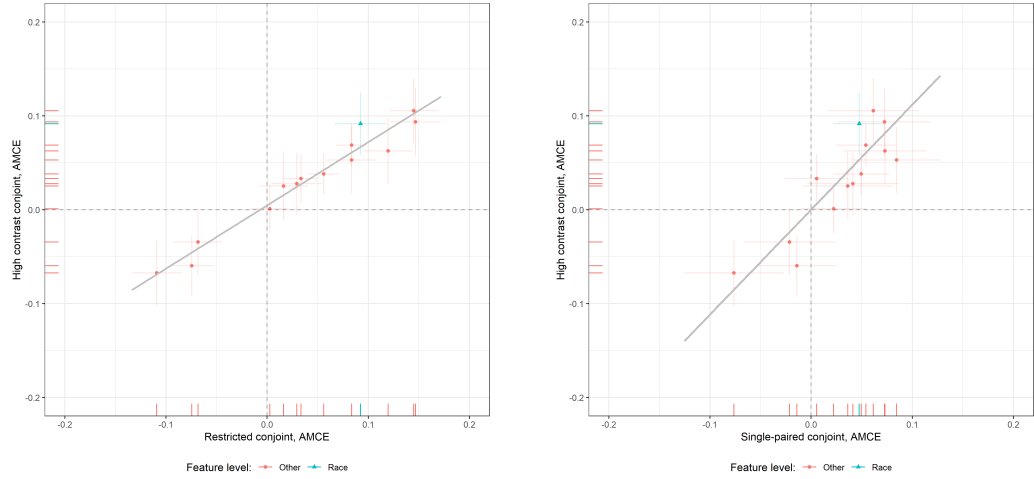
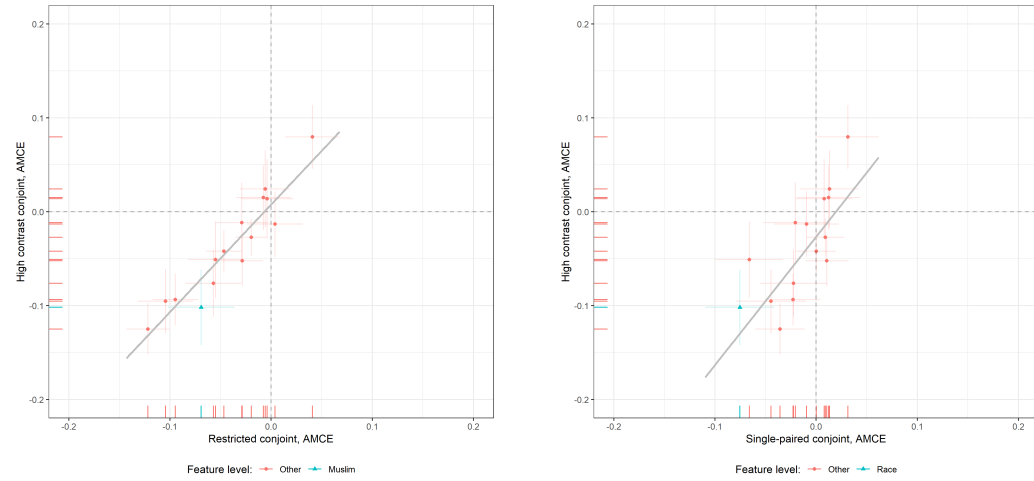


Figure 8. Study inspired by Hainmueller and Hopkins (2015). High-contrast estimates versus estimates from the alternative designs



Note: The figures show each AMCE-estimate from the high-contrast design versus estimates obtained from the restricted (left) and the single-profile designs (right). Each point represents an AMCE-estimate with 95 percent confidence intervals with the coefficients ordered by their magnitude from most negative to most positive. Sensitive features coloured blue. $N = 7059$

Limitations

It is worth acknowledging the limitations of the research articles contained within this dissertation. While each article includes a detailed discussion of the limitations of the research described therein, this section draws some general conclusions from those discussions and provides a few answers to some of these limitations.

First, several critiques have been levelled specifically at the correspondence studies on labour market discrimination. One question is whether we can infer real-life consequences such as labour market inequalities from the experiments (Fryer and Levitt 2004; Heckman 1998). For instance, the results of the field experiments in this dissertation measures average employer behaviour, but real applicants have information that guides their selection of jobs to which they will apply. Identifying prospective workplaces is not a random process. If job applicants are aware of potential discriminatory treatment by employers, for example, they may focus on jobs they perceive to be available to them. Fryer and Levitt (2004) have argued that the process of finding a job often works through unofficial channels like personal networks, suggesting that workers may often simply avoid discriminatory employers. Moreover, because the labour market is often segregated, minority and majority applicants might apply for different jobs (Heckman, Lyons, and Todd 2000). These endogenous behaviours are likely to lead to smaller differences than what is indicated by the average treatment effects measured in the field experiments. While these points of critique are important, they might not be as relevant to the Danish setting as they are to an American setting. The large public sector in Denmark relies on a formalized hiring process, which means that fewer jobs are distributed through informal networks. Moreover, the labour market in Denmark (and most European contexts) is significantly less segregated than the labour market in the US (Musterd 2005).

Second, some caveats apply to the generalizability of my results insofar as I relied heavily on empirical data from the Danish setting. Four of the research articles (A-D) unfold in the Danish context. However, the findings from the research articles generally align with the literature, suggesting that theories developed in very different contexts – often in an American setting – can be replicated in the Danish context. The replication of existing research is important to the

advancement of the social sciences and the replication in different contexts helps to establish the scope of the findings (Gerring 2011). The research articles thus contribute to the field by testing the extent to which established theoretical propositions developed in other contexts can be replicated in Denmark.

Third, the articles are limited in their capacity to empirically tease apart the theoretical mechanisms underpinning the results. In Chapter Two, I described the difficulties of distinguishing between taste-based and statistical discrimination. Because internal motivations are difficult to measure empirically (Pager and Shepherd 2008) or because both mechanisms exist simultaneously (Midtbøen 2013), evidence of discrimination often aligns with the theoretical implications of more than one explanation. I also proposed a conceptual division of statistical theories of discrimination into accurate (based on correct beliefs) and inaccurate (based on incorrect beliefs) types. While this refines the traditional distinctions applied to the concept of discrimination, it also complicates the tests needed to establish the underlying mechanisms. Although the research articles qualify the reasons underlying discriminatory behaviour, they cannot fully disentangle the importance of taste-based discrimination vis-à-vis accurate or inaccurate statistical discrimination.

Fourth, the articles focus on the largest minority group in Denmark, immigrant-origin minorities of Middle Eastern descent, and the results cannot be straightforwardly transferred to other immigrant-origin groups. The research articles demonstrate that discrimination depends on combinations of attributes, and on the convergence and the salience of these attributes. In other words, the results do not necessarily speak to potential discrimination against other immigrant-origin groups. In the same vein, the studies rely on a large pool of the names most commonly used by immigrant-origin minorities in Denmark. While this choice was made in order to increase the studies' real-world relevance, the most commonly used names are also likely to trigger ethnic stereotypes. It might be the case that less common immigrant-origin names result in smaller ethnic penalties.

Implications

In the five research articles, I argue and present evidence showing that simple group-category cues can sufficiently and deeply affect human attitudes and behaviours motivating group-based biases. Taken all together, the studies contribute to a larger conversation concerning how group-based biases shape the prospects of immigrant-origin minorities in Western societies.

A rather discouraging finding from this dissertation is that discrimination appears difficult to mitigate and immigrant-origin minorities have few tools at their disposal to reduce discrimination. All else being equal, immigrant-origin minorities face discriminatory behaviour irrespective of whether they are well-educated and have labour market experience (article A), indicate that they have adopted cultural norms in their job applications (article B), counter beliefs that they are less likely to vote when they approach politicians for constituency services (article C), or when they run for political office despite the fact that they hold political experience, education and policy positions to the same extent as other political candidates (article D).

Overall, these results suggest that discrimination feeds into what Adida, Laitin, and Valfort (2016) define as a ‘self-reinforcing discrimination equilibrium’. On the one hand, discrimination faced by minorities is likely to cause them to stay more sequestered in their own communities and norms, which in turn aids majorities in their efforts to rationalize discrimination because they observe confirmations of their stereotypes. On the other hand, immigrant-origin minorities face discrimination irrespective of their behaviour. This creates a self-reinforcing cycle that is hard to break and ultimately may hinder integration.

The findings hold implications for the literature on the cumulative effects of discrimination (e.g. Oskooii (2018) and Pager and Shepherd (2008)). Specifically, the effects of discrimination may span domains and extend forward in time, magnifying their initial impact. While the research articles in this dissertation focused on individual decision points, the results highlight that discrimination is a factor in a multitude of settings. Indeed, the effects are likely to expand or even mutually reinforce one another across empirical settings (Blank 2005; Adida, Laitin, and Valfort 2016; Pager and Shepherd 2008). For example, recent studies have indicated that experiences of discrimination

affect levels of social trust and political engagement and have been documented as negatively affecting citizens' trust in government and their perceptions of fairness (Oskooii 2018; Wong et al. 2011). At the broader societal level, empirical evidence has shown that discrimination is not only economically inefficient and illegal, it is also a source of social unrest: when immigrant-origin minority groups are disadvantaged in their access to economic resources and political influence, it can provoke violent conflicts (Dancygier 2010; Dancygier and Laitin 2014).

Showing that different groups face biases calls for policy interventions to reduce the extent and impact of discriminatory behaviour, thus it is worth considering some institutional implications of the findings. Articles A and B indicate that, at the organisational level, employers seem to face a problem abiding with the law regarding equal treatment of job applicants. This is a problem that must be tackled, not least within the large public sector that hires a significant share of the Danish workforce. It calls for more work to explore bias-reducing mechanisms and interventions – an important gap in the research on discrimination (Butler and Crabtree 2017). A recent study from the Danish context has shown that it *is* possible to implement simple tweaks to minimize the extent and impact of discrimination. Andersen and Guul (2018) demonstrate that cognitive pressures and workload affect discriminatory behaviour and that allowing HR-personnel sufficient time to read applications is a simple measure to target discrimination. Other tactics at the organizational level could be diversity trainings that focus on breaking down prejudice among natives or the use of anonymous job applications. Building on the findings in this dissertation, workplaces should invest in developing more efficient de-biasing strategies.

In terms of the findings on how group categories shape political representation, the articles hold some important implications as well. Communication between constituents and local legislators is an important avenue for voters to voice their political concerns; this type of communication matters for how legislators make policy decisions. Moreover, some evidence suggests that when underrepresented voter groups view their representatives as more responsive, they are more likely to participate in politics (Chattopadhyay and Duflo 2004; Griffin and Keane 2006). The inadequate political responsiveness to immigrant-origin minorities indicates that an already disadvantaged group is further constrained in their ability to influence and participate in the political system. Beyond its

academic interest, the motivations underlying ethnocentric behaviour tap into fundamental questions about the potential for promoting political equality through political-institutional design. In research article C, *Who is responsive*, we find compelling evidence indicating that minority voters can identify representatives who are more responsive to them by paying attention to the candidate's party *and* the candidate's personal political preferences. Accordingly, this finding highlights an appealing feature of the open-list system, where personal votes matter more. The personal vote enhances the possibility for voters to institute political accountability and vote for candidates who are significantly less likely to discriminate than the average politician.

Article D, *Candidate choice*, presents novel evidence relevant to questions concerning the underrepresentation of immigrant-origin minorities in political legislatures. Clearly, any given social group will tend to be underrepresented in legislatures if they are less likely to be qualified, have resources and run for office (Carnes and Lupu 2016). Our findings suggest that voter biases also contribute to immigrant-origin candidates' numerical underrepresentation. These results have a bearing on discussions about mechanisms for encouraging the election of immigrant-origin minorities. One mechanism to improve representation that appears particularly feasible is to increase the turnout among immigrant-origin voters. Field experiments conducted in Denmark indicate that it is possible to mobilize young low-propensity voters by sending them mobilization letters prior to elections (Bhatti et al. 2015). Further, once some members of the household vote, this has a positive spill-over effect on other members of the household (Dahlgaard 2018). The findings also have a bearing on political recruitment processes to address disparities in representation. Specifically, party gatekeepers have control over the recruitment and placement of political candidates on ballots, which can be instrumental in electing a larger share of immigrant-origin minorities (Carnes 2016; Fox and Lawless 2010; Carroll and Sanbonmatsu 2013).

Future research

The above-mentioned implications suggest several questions ripe for further inquiry. Future research should take place on at least three fronts.

First, this dissertation has outlined a problem with the traditional way of conceiving discrimination through the lens of the two workhorse models: taste-based and statistical discrimination. As outlined previously, actors may have inaccurate beliefs about the behavioural patterns of a social group. This conceptual distinction is rarely discussed, but it is important in order to identify and interpret the motives and implications of discriminatory behaviour. Despite the fact that discrimination has been widely studied in the social sciences, conceptual work on the sources of discrimination – as well as advances in empirical testing – is important to bring the field forward. Specifically, further integration of traditional economic models of discrimination and research on psychological biases and heuristics for inaccurate stereotype formation is an important area for future work to consider. This distinction is not only important from an academic perspective. As research on discrimination advances in identifying driving factors for discrimination, it will support the development of more effective policy interventions.

Second, the results in this dissertation point to the need for considering proposals on how to mitigate discriminatory behaviour. What can be done to reduce the bias uncovered? As briefly outlined above, advancing our knowledge on what can be done at the organizational level is important. This may involve collaborations between researchers and organizations willing to experiment with how they recruit future employers. If public sector workplaces (and workplaces in general) are serious about upholding meritocratic principles, they should apply experimental studies in order to test the effectiveness of various interventions.

Another venue for research on how discrimination can be mitigated is more research on which strategies individuals can follow to improve their outcomes. The results presented in this dissertation indicates that including subtle information in job applications that signal adaption to cultural norms is not a viable strategy, but that visible cues of religious practice matter. In the same vein, a correspondence experiment conducted in France, Duguet et al. (2010) finds that discrimination

of applicants with Moroccan and first and surname is significantly reduced when applicants have a French first name and a Moroccan surname. This suggests that adopting a 'dual name structure' using generic names might be an effective strategy for immigrant-origin minorities to avoid discrimination. Although it is undeniably a cost to discard a traditional name or avoid wearing visible religious symbols, it shows how employers respond to signals over which immigrant-origin minorities have some control. Future research should explore which tools immigrant-origin minorities have at their disposal to escape differential treatment.

Finally, studying the effects of interventions that focus on how information that renders discrimination objective and visible is an interesting avenue for future research. As pointed out by Adida, Laitin, and Valfort (2016), individuals are prone to change their behaviour when it indicates that institutional norms are violated. Documenting and informing the public on the findings on discrimination might change attitudes (Stoker 1998; Broockman and Kalla 2016) and behaviour (Pope, Price, and Wolfers 2018) (although see Butler and Crabtree (2017) for a discouraging null-effect of providing information on reducing bias among public officials). Pope, Price, and Wolfers (2018) demonstrate that widespread media attention following an academic study on bias among professional basketball referees led the bias to disappear. On a micro-level, recent studies indicate that perspective-taking exercises can persuade citizens to adopt more inclusionary behaviour (Adida, Lo, and Platas 2018) and these effects can be long-lasting (Broockman and Kalla 2016).

In summary, this dissertation has provided hard evidence on discrimination against immigrant-origin minority citizens indicating that differential treatment decisively influence their economic and political opportunities. The aggregate consequence of this behaviour is that immigrant-origin groups are at a disadvantage in the labour market as well as in the political system. The quest for future research is to develop better tests to identify reasons for discriminatory behaviour and, in a related vein, identifying effective and sustainable means to mitigate discriminatory behaviour.

Summary

In this dissertation, I explore how group-based biases shape economic and political interactions between salient social groups. Specifically, I test if, when and how some individuals are treated differently because of their descriptive characteristics such as ethnicity or gender. I employ a series of experiments to uncover these questions.

I apply a theoretical framework asserting that discrimination can be due to both personal preferences and strategic behaviour and draw upon insights from political behaviour and social psychology to better understand the theoretical underpinnings of discrimination. Specifically, I incorporate insights from a social cognition perspective, which offers a way to understand the cognitive processes by which people place others into social groups and how this shapes behaviour. From these perspectives, I lay out some propositions that I test in two empirical tracks across five research articles that all build on field or survey experiments.

In the first track, I explore how social group categories shape citizens' encounters with public managers and private employers during the hiring process in the Danish labour market. In two correspondence experiments in which equivalent job applications and cover letters with randomly assigned aliases were sent in response to job openings, I uncover differential treatment in hiring decisions. The experiments leave no doubt that immigrant-origin minorities are targets of significant discrimination. This differential treatment is startling considering the fact that applicants were highly qualified for the jobs they applied for. Going beyond existing work, I show that this is especially true when minorities are male or when female applicants wear a headscarf which suggests the importance of the intersection of ethnicity, gender and cues of cultural distinctiveness. Moreover, I find little evidence to indicate that immigrant-origin minorities can reduce this discrimination by indicating adherence to cultural norms.

In the second track, I study the effect of group-based biases on the political representation of underrepresented groups. The research articles present compelling evidence that immigrant-origin minorities face significant barriers in obtaining substantive and descriptive political representation. In a field experiment, the third research article indicates the significant bias of incumbents

in their direct communication with ethnic out-group constituents. This manifests itself directly in the legislator-constituent relationship: when constituents contact their local incumbents to retrieve information on the location of their polling station, minority voters are significantly less likely to receive a reply, and they receive replies of lower quality. Although the overall level of responsiveness increases when politicians face strong electoral incentives, the bias persists. One important contribution is the discovery that immigrant-origin voters can identify more responsive politicians by paying attention to two types of heuristics regarding legislators: their partisan affiliation cues and their stated preferences on immigration policies.

Departing from the finding that descriptive representation impacts substantive representation, the fourth research article explores reasons for the gap in political representation. Specifically, it investigates whether local political candidates with immigrant-origin names face barriers due to negative voter preferences. Building on a conjoint experiment, the article presents evidence indicating that the electoral prospects of political candidates with immigrant-origin names are hampered because voters prefer ethnic in-group candidates. Strikingly, this is true in a high-information setting where voters are informed about candidates' political experience, policy positions and party membership. Moreover, there is no evidence for a pro-male bias. Finally, in the last research article, I study the validity of the candidate conjoint experimental design. Specifically, I examine to what extent social desirability bias threatens validity and which tactics researchers can pursue to obtain reliable answers. The results indicate that social desirability bias may be a more minimal concern than what is often assumed.

Taken together, the evidence from the five research articles provides insight into a deeply challenging social issue. There are often strong legal or normative arguments emphasizing why, in many socio-political interactions, individuals' immutable group categories should be invisible. Inadequate representation and opportunities can have serious consequences and downstream electoral effects on a number of societal outcomes and have negative spill-over effects across social domains and time. The research articles indicate that discrimination appears to be hard to mitigate and immigrant-origin minorities have few tools at their disposal to reduce discrimination, which points to the need for institutional actions to eliminate barriers that inhibit individuals from attaining equal access.

References

- Aarøe, Lene, Michael Bang Petersen, and Kevin Arceneaux (2017). “The behavioral immune system shapes political intuitions: Why and how individual differences in disgust sensitivity underlie opposition to immigration”. In: *American Political Science Review* 111.2, pp. 277–294.
- Adida, Claire L, David D Laitin, and Marie-Anne Valfort (2016). *Why Muslim integration fails in Christian-heritage societies*. Harvard University Press.
- Adida, Claire L, Adeline Lo, and Melina R Platas (2018). “Perspective taking can promote short-term inclusionary behavior toward Syrian refugees”. In: *Proceedings of the National Academy of Sciences* 115.38, pp. 9521–9526.
- Aigner, Dennis J and Cain (1977). “Statistical theories of discrimination in labor markets”. In: *ILR Review* 30.2, pp. 175–187.
- Akerlof, George (1976). “The economics of caste and of the rat race and other woeful tales”. In: *The Quarterly Journal of Economics*, pp. 599–617.
- Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew (1954). “The nature of prejudice”. In:
- Altonji, Joseph G and Rebecca M Blank (1999). “Race and gender in the labor market”. In: *Handbook of labor economics* 3, pp. 3143–3259.
- Andersen, Simon Calmar and Thorbjørn Sejr Guul (2018). “Reducing Minority Discrimination at the Front Line—Combined Survey and Field Experimental Evidence”. In: *Journal of Public Administration Research and Theory*.
- Arrow, Kenneth et al. (1973). “The theory of discrimination”. In: *Discrimination in labor markets* 3.10, pp. 3–33.
- Athey, Susan and Guido W Imbens (2017). “The econometrics of randomized experiments”. In: *Handbook of Economic Field Experiments*. Vol. 1. Elsevier, pp. 73–140.
- Ayres, Ian and Peter Siegelman (1995). “Race and gender discrimination in bargaining for a new car”. In: *The American Economic Review*, pp. 304–321.

References

- Baars, Bernard J et al. (1997). *In the theater of consciousness: The workspace of the mind*. Oxford University Press, USA.
- Banducci, Susan A. et al. (2008). “Ballot photographs as cues in low-information elections”. In: *Political psychology* 29.6, pp. 903–917.
- Banton, Michael (1997). “The ethics of practice-testing”. In: *Journal of Ethnic and Migration Studies*.
- Barabas, Jason and Jennifer Jerit (2010). “Are survey experiments externally valid?” In: *American Political Science Review* 104.2, pp. 226–242.
- Beal, Frances M (2008). “Double jeopardy: To be Black and female”. In: *Meridians: feminism, race, transnationalism* 8.2, pp. 166–176.
- Becker, Gary S (1957). *The economics of discrimination*. University of Chicago press.
- Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan (2005). “Implicit discrimination”. In: *American Economic Review* 95.2, pp. 94–98.
- Bertrand, Marianne and Esther Duflo (2017). “Field experiments on discrimination”. In: *Handbook of Economic Field Experiments*. Vol. 1. Elsevier, pp. 309–393.
- Bertrand, Marianne and Sendhil Mullainathan (2004). “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination”. In: *American economic review* 94.4, pp. 991–1013.
- Bhatti, Yosef et al. (2015). “Getting out the vote with evaluative thinking”. In: *American Journal of Evaluation* 36.3, pp. 389–400.
- Blank, Rebecca M (2005). “Tracing the economic impact of cumulative discrimination”. In: *American Economic Review* 95.2, pp. 99–103.
- Bloemraad, Irene (2013). “Accessing the corridors of power: Puzzles and pathways to understanding minority representation”. In: *West European Politics* 36.3, pp. 652–670.
- Blumer, Herbert (1958). “Race prejudice as a sense of group position”. In: *Pacific sociological review* 1.1, pp. 3–7.
- Bodenhausen, Galen V (1990). “Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination”. In: *Psychological Science* 1.5, pp. 319–322.

- Bohren, J Aislinn et al. (2019). “Inaccurate Statistical Discrimination”. In: *NBER*.
- Bordalo, Pedro et al. (2016). “Stereotypes”. In: *The Quarterly Journal of Economics* 131.4, pp. 1753–1794.
- Broockman (2013). “Black politicians are more intrinsically motivated to advance blacks’ interests: A field experiment manipulating political incentives”. In: *American Journal of Political Science* 57.3, pp. 521–536.
- Broockman and Joshua Kalla (2016). “Durably reducing transphobia: A field experiment on door-to-door canvassing”. In: *Science* 352.6282, pp. 220–224.
- Broockman and Soltas (2019). *A natural experiment on taste-based racial and ethnic discrimination in elections*.
- Butler (2014). *Representing the advantaged: How politicians reinforce inequality*. Cambridge University Press.
- Butler and Broockman (2011). “Do politicians racially discriminate against constituents? A field experiment on state legislators”. In: *American Journal of Political Science* 55.3, pp. 463–477.
- Butler and Charles Crabtree (2017). “Moving beyond measurement: Adapting audit studies to test bias-reducing interventions”. In: *Journal of Experimental Political Science* 4.1, pp. 57–67.
- Butler and Jonathan Homola (2017). “An empirical justification for the use of racially distinctive names to signal race in experiments”. In: *Political Analysis* 25.1, pp. 122–130.
- Butler, Christopher F Karpowitz, and Pope (2012). “A field experiment on legislators’ home styles: service versus policy”. In: *The Journal of Politics* 74.2, pp. 474–486.
- Cain, John Ferejohn, and Morris Fiorina (1987). *The personal vote: Constituency service and electoral independence*. Harvard University Press.
- Canan, Coskun and Naika Foroutan (2016). “Changing perceptions? Effects of multiple social categorisation on German population’s perception of Muslims”. In: *Journal of Ethnic and Migration Studies* 42.12, pp. 1905–1924.
- Carnes, Nicholas (2013). *White-collar government: The hidden role of class in economic policy making*. University of Chicago Press.
- (2016). *Keeping workers off the ballot*. Tech. rep. Working Paper.

References

- Carnes, Nicholas and Noam Lupu (2016). “Do voters dislike working-class candidates? Voter biases and the descriptive underrepresentation of the working class”. In: *American Political Science Review* 110.4, pp. 832–844.
- Carroll, Susan J. and Kira Sanbonmatsu (2013). *More women can run: Gender and pathways to the state legislatures*. Oxford University Press.
- Chaiken, Shelly and Yaacov Trope (1999). *Dual-process theories in social psychology*. Guilford Press.
- Chattopadhyay, Raghabendra and Esther Duflo (2004). “Women as policy makers: Evidence from a randomized policy experiment in India”. In: *Econometrica* 72.5, pp. 1409–1443.
- Christensen, Darin and Simon Ejdemyr (2017). “Do Elections Improve Constituency Responsiveness? Evidence from US Cities”. In: *Political Science Research and Methods*, pp. 1–18.
- Coppock, Alexander (2016). “Randomizr: Easy to use tools for common forms of random assignment and sampling”. In: *R package version 0.5. 0*.
- Costa, Mia (2017). “How responsive are political elites? A meta-analysis of experiments on public officials”. In: *Journal of Experimental Political Science* 4.3, pp. 241–254.
- Crisp, Richard J and Miles Hewstone (2007). *Multiple social categorization: Processes, models and applications*. Psychology Press.
- Crisp, Richard J, Miles Hewstone, and Mark Rubin (2001). “Does multiple categorization reduce intergroup bias?” In: *Personality and social psychology bulletin* 27.1, pp. 76–89.
- Dahlgard, Jens Olav (2018). “Trickle-up political socialization: The causal effect on turnout of parenting a newly enfranchised voter”. In: *American Political Science Review* 112.3, pp. 698–705.
- Danckert, Bolette, Peter Thisted Dinesen, and Kim Mannemar Sønderskov (2017). “Reacting to neighborhood cues? Political sophistication moderates the effect of exposure to immigrants”. In: *Public opinion quarterly* 81.1, pp. 37–56.
- Dancygier (2010). *Immigration and conflict in Europe*. Cambridge University Press.
- Dancygier and David D Laitin (2014). “Immigration into Europe: Economic discrimination, violence, and public policy”. In: *Annual Review of Political Science* 17, pp. 43–64.

- Desposato, Scott (2015). *Ethics and experiments: Problems and solutions for social scientists and policy professionals*. Routledge.
- Devine, Patricia G (1989). "Stereotypes and prejudice: Their automatic and controlled components." In: *Journal of personality and social psychology* 56.1, p. 5.
- Dovidio, John F (2010). *The SAGE handbook of prejudice, stereotyping and discrimination*. Sage Publications.
- Druckman, James N, Thomas J Leeper, and Kevin J Mullinix (2014). "The experimental study of legislative behaviour". In: *The Oxford Handbook of Legislative Studies*. Oxford Handbooks in Politics &, p. 194.
- Druckman, James N and Arthur Lupia (2012). "Experimenting with politics". In: *Science* 335.6073, pp. 1177–1179.
- Duguet, Emmanuel et al. (2010). "Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area". In: *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, pp. 187–215.
- Eagly, Alice H and Shelly Chaiken (1998). "Attitude structure and function." In:
- Eagly, Alice H and Mary E Kite (1987). "Are stereotypes of nationalities applied to both women and men?" In: *Journal of personality and social psychology* 53.3, p. 451.
- Edelman, Benjamin, Michael Luca, and Dan Svirsky (2017). "Racial discrimination in the sharing economy: Evidence from a field experiment". In: *American Economic Journal: Applied Economics* 9.2, pp. 1–22.
- Enos, Ryan D (2017). *The space between us: Social geography and politics*. Cambridge University Press.
- Epp, Charles R, Steven Maynard-Moody, and Donald Haider-Markel (2017). "Beyond profiling: The institutional sources of racial disparities in policing". In: *Public Administration Review* 77.2, pp. 168–178.
- Fenno, Richard F (1978). *Home style: House members in their districts*. HarperCollins,
- Fisher et al. (2015). "Candidate ethnicity and vote choice in Britain". In: *British Journal of Political Science* 45.4, pp. 883–905.

References

- Fiske, Susan T (1998). “Stereotyping, prejudice, and discrimination”. In: *The handbook of social psychology* 2, pp. 357–411.
- Fiske, Susan T and Steven L Neuberg (1990). “A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation”. In: *Advances in experimental social psychology*. Vol. 23. Elsevier, pp. 1–74.
- Fox, Richard L. and Jennifer L. Lawless (2010). *It still takes a candidate: Why women don’t run for office*. Cambridge University Press.
- Fryer, Roland G and Steven D Levitt (2004). “The causes and consequences of distinctively black names”. In: *The Quarterly Journal of Economics* 119.3, pp. 767–805.
- Gaddis, S Michael (2017). “How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies”. In: *Sociological Science* 4, pp. 469–489.
- Gaertner, Samuel L and John F Dovidio (2005). “Categorization, recategorization, and intergroup bias”. In: *On the nature of prejudice: Fifty years after Allport*, pp. 71–88.
- Gell-Redman, Micah et al. (2018). “It’s all about race: How state legislators respond to immigrant constituents”. In: *Political Research Quarterly* 71.3, pp. 517–531.
- Gerber, Alan S and Donald P Green (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.
- Gerring, John (2011). *Social science methodology: A unified framework*. Cambridge University Press.
- Gilens, Martin (1996). ““Race coding” and white opposition to welfare”. In: *American Political Science Review* 90.3, pp. 593–604.
- Gnambs, Timo and Kai Kaspar (2017). “Socially desirable responding in web-based questionnaires: A meta-analytic review of the candor hypothesis”. In: *Assessment* 24.6, pp. 746–762.
- Griffin, John D and Michael Keane (2006). “Descriptive representation and the composition of African American turnout”. In: *American Journal of Political Science* 50.4, pp. 998–1012.
- Grose (2011). *Congress in black and white: Race and representation in Washington and at home*. Cambridge University Press.
- (2014). “Field experimental work on political institutions”. In: *Annual Review of Political Science* 17, pp. 355–370.

-
- (2016). “Field experiments on elected and public officials: Ethical obligations and requirements”. In: *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, pp. 227–38.
- Guryan, Jonathan and Kerwin Kofi Charles (2013). “Taste-based or statistical discrimination: the economics of discrimination returns to its roots”. In: *The Economic Journal* 123.572, F417–F432.
- Hainmueller, Jens and Dominik Hangartner (2013). “Who gets a Swiss passport? A natural experiment in immigrant discrimination”. In: *American political science review* 107.1, pp. 159–187.
- Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto (2014). “Do survey experiments capture real-world behavior? External validation of conjoint and vignette analyses with a natural experiment”. In: *Proceedings of the National Academy of Sciences* 112.8.
- (2015). “Validating vignette and conjoint survey experiments against real-world behavior”. In: *Proceedings of the National Academy of Sciences* 112.8, pp. 2395–2400.
- Hainmueller, Jens and Daniel J Hopkins (2014). “Public attitudes toward immigration”. In: *Annual Review of Political Science* 17, pp. 225–249.
- (2015). “The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants”. In: *American Journal of Political Science* 59.3, pp. 529–548.
- Hainmueller, Jens, Daniel J Hopkins, and Teppei Yamamoto (2014). “Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments”. In: *Political Analysis* 22.1, pp. 1–30.
- Hariri, Jacob Gerner and David Dreyer Lassen (2017). “Income and outcomes: Social desirability bias distorts measurements of the relationship between income and political behavior”. In: *Public Opinion Quarterly* 81.2, pp. 564–576.
- Harnois, Catherine E (2015). “Jeopardy, consciousness, and multiple discrimination: intersecting inequalities in contemporary Western Europe”. In: *Sociological Forum*. Vol. 30. 4. Wiley Online Library, pp. 971–994.

References

- Hastie, Reid and Purohit A Kumar (1979). “Person memory: Personality traits as organizing principles in memory for behaviors.” In: *Journal of Personality and Social Psychology* 37.1, p. 25.
- Heckman, James J (1998). “Detecting discrimination”. In: *Journal of economic perspectives* 12.2, pp. 101–116.
- Heckman, James J, Thomas M Lyons, and Petra E Todd (2000). “Understanding black-white wage differentials, 1960-1990”. In: *American Economic Review* 90.2, pp. 344–349.
- Hedegaard, Morten Størting and Jean-Robert Tyran (2018). “The price of prejudice”. In: *American Economic Journal: Applied Economics* 10.1, pp. 40–63.
- Hemker, Johannes and Anselm Rink (2017). “Multiple dimensions of bureaucratic discrimination: Evidence from German welfare offices”. In: *American Journal of Political Science* 61.4, pp. 786–803.
- Hilton, James L and William Von Hippel (1996). “Stereotypes”. In: *Annual review of psychology* 47.1, pp. 237–271.
- Hjorth, Frederik (2017). “The influence of local ethnic diversity on group-centric crime attitudes”. In: *British Journal of Political Science*, pp. 1–23.
- Holland, Paul W (1986). “Statistics and causal inference”. In: *Journal of the American statistical Association* 81.396, pp. 945–960.
- Horiuchi, Yusaku, Daniel M Smith, and Teppei Yamamoto (2018). “Identifying voter preferences for politicians’ personal attributes: A conjoint experiment in Japan”. In: *Political Science Research and Methods*, pp. 1–17.
- Iyer, Lakshmi et al. (2012). “The power of political voice: women’s political representation and crime in India”. In: *American Economic Journal: Applied Economics* 4.4, pp. 165–93.
- Janus, Alexander L (2010). “The influence of social desirability pressures on expressed immigration attitudes”. In: *Social Science Quarterly* 91.4, pp. 928–946.
- Janusz, Andrew and Nazita Lajevardi (2016). “The political marginalization of Latinos: Evidence from three field experiments”. In: *Available at SSRN 2799043*.
- Kaas, Leo and Christian Manger (2012). “Ethnic discrimination in Germany’s labour market: a field experiment”. In: *German economic review* 13.1, pp. 1–20.

- Kahneman, Daniel et al. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.
- Kaminska, Olena and Tom Foulsham (2013). *Understanding sources of social desirability bias in different modes: Evidence from eye-tracking*. Tech. rep. ISER Working Paper Series.
- Kinder, Donald R and Cindy D Kam (2010). *Us against them: Ethnocentric foundations of American opinion*. University of Chicago Press.
- Kinder, Donald R and David O Sears (1981). “Prejudice and politics: Symbolic racism versus racial threats to the good life.” In: *Journal of personality and social psychology* 40.3, p. 414.
- Kirkland, Patricia A and Alexander Coppock (2017). “Candidate Choice Without Party Labels”. In: *Political Behavior*, pp. 1–21.
- Kline, Patrick (2011). “Oaxaca-Blinder as a reweighting estimator”. In: *American Economic Review* 101.3, pp. 532–37.
- Knox, Dean, Will Lowe, and Jonathan Mummolo (2019). “The Bias Is Built In: How Administrative Records Mask Racially Biased Policing”. In: *Available at SSRN*.
- Koopmans, Ruud (2016). “Does assimilation work? Sociocultural determinants of labour market participation of European Muslims”. In: *Journal of Ethnic and Migration Studies* 42.2, pp. 197–216.
- Krumpal, Ivar (2013). “Determinants of social desirability bias in sensitive surveys: a literature review”. In: *Quality & Quantity* 47.4, pp. 2025–2047.
- Kuklinski, James H et al. (1997). “Racial prejudice and attitudes toward affirmative action”. In: *American Journal of Political Science*, pp. 402–419.
- Kuo, Alexander, Neil Malhotra, and Cecilia Hyunjung Mo (2017). “Social exclusion and political identity: The case of Asian American partisanship”. In: *The Journal of Politics* 79.1, pp. 17–32.
- Lasswell, Harold Dwight (1950). *Politics: Who gets what, when, how*. P. Smith New York.
- Lippert-Rasmussen, Kasper (2014). *Born free and equal?: a philosophical inquiry into the nature of discrimination*. Oxford University Press.
- Lipsky, Michael (1980). “Street-Level Bureaucracy: The Dilemmas of the Individual in Public Service: Russell Sage Foundation”. In: *New York*.

References

- Lupia, Arthur (1994). “Shortcuts versus encyclopedias: Information and voting behavior in California insurance reform elections”. In: *American Political Science Review* 88.1, pp. 63–76.
- Ma, Debbie S et al. (2013). “When fatigue turns deadly: The association between fatigue and racial bias in the decision to shoot”. In: *Basic and applied social psychology* 35.6, pp. 515–524.
- Macrae, C Neil and Galen V Bodenhausen (2000). “Social cognition: Thinking categorically about others”. In: *Annual review of psychology* 51.1, pp. 93–120.
- Mansbridge, Jane (1999). “Should blacks represent blacks and women represent women? A contingent” yes””. In: *The Journal of politics* 61.3, pp. 628–657.
- Mason, Lilliana (2018). *Uncivil agreement: How politics became our identity*. University of Chicago Press.
- McClendon, Gwyneth H (2016). “Race and responsiveness: An experiment with South African politicians”. In: *Journal of Experimental Political Science* 3.1, pp. 60–74.
- McDermott, Monika L (1998). “Race and gender cues in low-information elections”. In: *Political Research Quarterly* 51.4, pp. 895–918.
- Mendez, Matthew and Grose (2018). “Doubling down: Inequality in responsiveness and the policy preferences of elected officials”. In: *Legislative Studies Quarterly* 43.3, pp. 457–491.
- Midtbøen (2013). “Determining discrimination. A multi-method study of employment discrimination among descendants of immigrants in Norway.” In:
- (2015). “The context of employment discrimination: interpreting the findings of a field experiment”. In: *The British journal of sociology* 66.1, pp. 193–214.
- (2016). “Discrimination of the second generation: Evidence from a field experiment in Norway”. In: *Journal of International Migration and Integration* 17.1, pp. 253–272.
- Mo, Cecilia Hyunjung (2015). “The consequences of explicit and implicit gender attitudes and candidate quality in the calculations of voters”. In: *Political Behavior* 37.2, pp. 357–395.
- Moulton, Brent R (1990). “A reexamination of the federal-private wage differential in the United States”. In: *Journal of Labor Economics* 8.2, pp. 270–293.
- Musterd, Sako (2005). “Social and ethnic segregation in Europe: Levels, causes, and effects”. In: *Journal of urban affairs* 27.3, pp. 331–348.

- Nunley, John M et al. (2016). “College major, internship experience, and employment opportunities: Estimates from a résumé audit”. In: *Labour Economics* 38, pp. 37–46.
- Nye, John V.C., Ilia Rainer, and Thomas Stratmann (2014). “Do black mayors improve black relative to white employment outcomes? Evidence from large US cities”. In: *The Journal of Law, Economics, & Organization* 31.2, pp. 383–430.
- Ono, Yoshikuni and Masahiro Yamada (2016). “Do Voters Prefer Gender Stereotypic Candidates?: Evidence from a Conjoint Survey Experiment in Japan”. In:
- Oreopoulos, Philip (2011). “Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes”. In: *American Economic Journal: Economic Policy* 3.4, pp. 148–71.
- Oskooii, Kassra AR (2018). “Perceived Discrimination and Political Behavior”. In: *British Journal of Political Science*, pp. 1–26.
- Ostfeld, Mara (2017). “The backyard politics of attitudes toward immigration”. In: *Political Psychology* 38.1, pp. 21–37.
- Pager, Devah (2008). *Marked: Race, crime, and finding work in an era of mass incarceration*. University of Chicago Press.
- Pager, Devah and Diana Karafin (2009). “Bayesian bigot? Statistical discrimination, stereotypes, and employer decision making”. In: *The Annals of the American Academy of Political and Social Science* 621.1, pp. 70–93.
- Pager, Devah and Lincoln Quillian (2005). “Walking the talk? What employers say versus what they do”. In: *American Sociological Review* 70.3, pp. 355–380.
- Pager, Devah and Hana Shepherd (2008). “The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets”. In: *Annu. Rev. Sociol* 34, pp. 181–209.
- Pager, Devah and Bruce Western (2012). “Identifying discrimination at work: The use of field experiments”. In: *Journal of Social Issues* 68.2, pp. 221–237.
- Paulhus, Delroy L (1984). “Two-component models of socially desirable responding.” In: *Journal of personality and social psychology* 46.3, p. 598.

References

- Pedersen, Mogens Jin, Justin M Stritch, and Frederik Thuesen (2018). "Punishment on the frontlines of public service delivery: Client ethnicity and caseworker sanctioning decisions in a Scandinavian welfare state". In: *Journal of Public Administration Research and Theory* 28.3, pp. 339–354.
- Phelps, Edmund S (1972). "The statistical theory of racism and sexism". In: *The american economic review*, pp. 659–661.
- Philpot, Tasha S and Hanes Walton (2007). "One of our own: Black female candidates and the voters who support them". In: *American Journal of Political Science* 51.1, pp. 49–62.
- Pope, Joseph Price, and Justin Wolfers (2018). "Awareness reduces racial bias". In: *Management Science* 64.11, pp. 4988–4995.
- Portmann, Lea and Nenad Stojanović (2019). "Electoral discrimination against immigrant-origin candidates". In: *Political Behavior* 41.1, pp. 105–134.
- Pratto, Felicia and John A Bargh (1991). "Stereotyping based on apparently individuating information: Trait and global components of sex stereotypes under attention overload". In: *Journal of Experimental Social Psychology* 27.1, pp. 26–47.
- Pratto, Felicia, Jim Sidanius, et al. (1994). "Social dominance orientation: A personality variable predicting social and political attitudes." In: *Journal of personality and social psychology* 67.4, p. 741.
- Putnam, Robert D (2007). "E pluribus unum: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture". In: *Scandinavian political studies* 30.2, pp. 137–174.
- Quillian, Lincoln et al. (2017). "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time". In: *Proceedings of the National Academy of Sciences* 114.41, pp. 10870–10875.
- Ransford, H Edward (1980). "The prediction of social behavior and attitudes". In: *Social stratification: A multiple hierarchy approach*, pp. 265–295.
- Riach, Peter A and Judith Rich (2004). "Deceptive field experiments of discrimination: are they ethical?" In: *Kyklos* 57.3, pp. 457–470.
- Rivera, Lauren A (2012). "Hiring as cultural matching: The case of elite professional service firms". In: *American sociological review* 77.6, pp. 999–1022.

- Sanchez, Gabriel R (2008). “Latino group consciousness and perceptions of commonality with African Americans”. In: *Social Science Quarterly* 89.2, pp. 428–444.
- Scheve, Kenneth F and Matthew J Slaughter (2001). “Labor market competition and individual preferences over immigration policy”. In: *Review of Economics and Statistics* 83.1, pp. 133–145.
- Schildkraut, Deborah J (2005). “The rise and fall of political engagement among Latinos: The role of identity and perceptions of discrimination”. In: *Political Behavior* 27.3, pp. 285–312.
- (2010). *Americanism in the twenty-first century: Public opinion in the age of immigration*. Cambridge University Press.
- Sears, David O (1986). “College sophomores in the laboratory: Influences of a narrow data base on social psychology’s view of human nature.” In: *Journal of personality and social psychology* 51.3, p. 515.
- Sen, Maya (2017). “How political signals affect public support for judicial nominations: Evidence from a conjoint experiment”. In: *Political Research Quarterly* 70.2, pp. 374–393.
- Sen, Maya and Omar Wasow (2016). “Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics”. In: *Annual Review of Political Science* 19, pp. 499–522.
- Sherman, Jeffrey W, C Neil Macrae, and Galen V Bodenhausen (2000). “Attention and stereotyping: Cognitive constraints on the construction of meaningful social impressions”. In: *European review of social psychology* 11.1, pp. 145–175.
- Sidanius, Jim and Felicia Pratto (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.
- Sniderman, Paul M, Louk Hagendoorn, and Markus Prior (2004). “Predisposing factors and situational triggers: Exclusionary reactions to immigrant minorities”. In: *American political science review* 98.1, pp. 35–49.
- Stigler, George J and Gary S Becker (1977). “De gustibus non est disputandum”. In: *The american economic review* 67.2, pp. 76–90.
- Stoker, Laura (1998). “Understanding whites’ resistance to affirmative action: The role of principled commitments and racial prejudice”. In: *Perception and prejudice: Race and politics in the United States*, pp. 135–70.

References

- Swain, Carol Miller (1995). *Black faces, black interests: The representation of African Americans in Congress*. Harvard University Press.
- Tajfel, Henri (1981). *Human groups and social categories: Studies in social psychology*. CUP Archive.
- (2010). *Social identity and intergroup relations*. Vol. 7. Cambridge University Press.
- Tajfel, Henri et al. (1971). “Social categorization and intergroup behaviour”. In: *European journal of social psychology* 1.2, pp. 149–178.
- Teele, Dawn Langan (2014). *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences*. Yale University Press.
- Teele, Dawn Langan, Joshua Kalla, and Frances Rosenbluth (2018). “The Ties That Double Bind: Social Roles and Women’s Underrepresentation in Politics”. In: *American Political Science Review* 112.3, pp. 525–541.
- Thrasher, Michael et al. (2017). “Candidate ethnic origins and voter preferences: Examining name discrimination in local elections in Britain”. In: *British Journal of Political Science* 47.2, pp. 413–435.
- Tjaden, Jasper Dag, Carsten Schwemmer, and Menusch Khadjavi (2018). “Ride with Me—Ethnic Discrimination, Social Markets, and the Sharing Economy”. In: *European Sociological Review* 34.4, pp. 418–432.
- Tourangeau, Roger and Ting Yan (2007). “Sensitive questions in surveys.” In: *Psychological bulletin* 133.5, p. 859.
- Turner, John C (1982). “Towards a cognitive redefinition of the social group”. In: *Social identity and intergroup relations*, pp. 15–40.
- Vernby, Kåre and Dancygier (2019). “Can immigrants counteract employer discrimination? A factorial field experiment reveals the immutability of ethnic hierarchies”. In: *PloS one* 14.7, e0218044.
- Villadsen, Anders R and Jesper N Wulff (2018). “Is the Public Sector a Fairer Employer? Ethnic employment discrimination in the public and private sectors”. In: *Academy of Management Discoveries* 4.4, pp. 429–448.

- Vries, Catherine E de, Elias Dinas, and Hector Solaz (2015). “You Have Got Mail! A Field Experiment on Legislator Responsiveness in the European Parliament”. In: *Paper to the Southern Political Science Association meeting, New Orleans, January*, pp. 15–17.
- Washington, Ebonya (2006). “How black candidates affect voter turnout”. In: *The Quarterly Journal of Economics* 121.3, pp. 973–998.
- Weber, Max (1978). *Economy and society: An outline of interpretive sociology*. Vol. 1. Univ of California Press.
- Wells, Gary L and Paul D Windschitl (1999). “Stimulus sampling and social psychological experimentation”. In: *Personality and Social Psychology Bulletin* 25.9, pp. 1115–1125.
- White, Ariel R, Noah L Nathan, and Julie K Faller (2015). “What do I need to vote? Bureaucratic discretion and discrimination by local election officials”. In: *American Political Science Review* 109.1, pp. 129–142.
- Wong, Janelle S et al. (2011). *Asian American political participation: Emerging constituents and their political identities*. Russell Sage Foundation.
- Wright, Matthew and Jack Citrin (2011). “Saved by the stars and stripes? Images of protest, salience of threat, and immigration attitudes”. In: *American Politics Research* 39.2, pp. 323–343.
- Wulff, Jesper N and Anders R Villadsen (2019). “Are Survey Experiments As Valid as Field Experiments in Management Research? An Empirical Comparison Using the Case of Ethnic Employment Discrimination”. In: *European Management Review*.
- Zschirnt, Eva (2016). “Revisiting ethics in correspondence testing”. In: *NCCR-on the move, working paper series* 8.
- Zschirnt, Eva and Didier Ruedin (2016). “Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015”. In: *Journal of Ethnic and Migration Studies* 42.7, pp. 1115–1134.

Experimental evidence of discrimination in the labour market: Intersections between ethnicity, gender, and socio-economic status

Malte Dahl and Niels Krog*

European Sociological Review (2018), Volume 34, Issue 4, Pages 402–417

Abstract

This article presents evidence of ethnic discrimination in the recruitment process from a field experiment conducted in the Danish labour market. In a correspondence experiment, fictitious job applications were randomly assigned either a Danish or Middle Eastern-sounding name and sent to real job openings. In addition to providing evidence on the extent of ethnic discrimination in the Danish labour market, the study offers two novel contributions to the literature more generally. First, because a majority of European correspondence experiments have relied solely on applications with male aliases, there is limited evidence on the way gender and ethnicity interact across different occupations. By randomly assigning gender and ethnicity, this study suggests that ethnic discrimination is strongly moderated by gender: minority males are consistently subject to a much larger degree of discrimination than minority females across different types of occupations. Second, this study addresses a key critique of previous correspondence experiments by examining the potential confounding effect of socio-economic status related to the names used to represent distinct ethnic groups. The results support the notion that differences in callbacks are caused exclusively by the ethnic traits.

*The authors would like to thank Peter T. Dinesen, Martin Vinaes Larsen and Jens Olav Dahlgaard for extremely valuable advices. The authors also thank participants at the annual meeting of the American Sociological Association, 2017 and participants at the annual meeting of the Danish Political Science Association, 2016 for comments on this article. The authors would also like to express their gratitude to the anonymous reviewers, who provided very helpful comments on an earlier version of this article. The replication data and code for conducting the data analysis are available at the authors' homepage.

Introduction

In many European countries, non-western immigrants face substantial employment deficits and wage differentials. This has grave consequences for the unemployed individuals and for the societies they inhabit, and has given rise to intense political debates about the rights and obligations of immigrant-origin minorities (Dancygier and Laitin 2014). There are a multitude of possible explanations for such labour market differentials, but several studies suggest that discrimination in the hiring process is an entry barrier to immigrant-origin minorities (Carlsson and Rooth 2012; Kaas and Manger 2012; Baert et al. 2015; Arai, Bursell, and Nekby 2016; Birkelund, Heggebø, and Rogstad 2017). A range of designs have been leveraged to study labour market discrimination, but field experiments are seen as a significant methodological advance (Bertrand and Duflo 2017; Neumark 2018). So-called *correspondence experiments*, where applications are submitted to job advertisements in the name of fictitious applicants, are an increasingly common tool for social science researchers. These experiments enable researchers to identify if, and to what extent, group affiliations—signified by names—affect the chances of getting a job interview.

Despite the fact that an increasing number of field experiments have examined labour market discrimination, important questions remain unanswered. First, in a European context, many correspondence studies have refrained from randomly assigning both ethnicity and gender, and therefore have overlooked the possible intersections between the two. By only using male applicants and assuming effect homogeneity across gender, many previous studies might not have told an accurate story about the overall ethnic disparities. If anything, the evidence from studies that do manipulate both traits points towards larger ethnic penalties among male applicants. Recent studies on gendered ethnic discrimination suggest that the interaction between gender and ethnicity is highly dependent on the composition of jobs included in the experiment, e.g. the share of private sector jobs or the gender composition in the occupations included (Bursell 2014; Midtbøen 2016). Hence, considering varying effect sizes across gender and across different occupations is essential to ensure generalizability when studying ethnic or racial discrimination. If such intersection between ethnicity and gender exists, it poses an important question of why members of the same ethnic group face

different outcomes.

Second, correspondence experiments face a challenge related to the internal validity when using names to manipulate characteristics of interest. Names do not exclusively signify racial or ethnic affiliation but contain a bundle of information. For example, the applicants' socio-economic status (SES) might be inferred from their names, and if the popular majority names used in correspondence experiments are perceived as having a higher SES than the distinct ethnic or racial minority names, it confounds the ethnic trait (Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004). In other words, if popular majority and immigrant-origin minority names are also a comparison across SES, it violates the excludability assumption that the effect is caused solely by the ethnic trait (Butler and Tavits 2017). Ultimately, this implies that it is not known whether ethnic or SES discrimination causes the ethnic disparities found in previous experiments.

To address these questions, we conducted a field experiment in which 800 generic applications were sent to job openings in the Danish labour market. Each job opening received two equally qualified applications, which were randomly assigned either a traditional Danish-sounding name or a Middle Eastern-sounding name. Applicants' ethnicity and gender were randomly assigned to study the interaction between applicant gender and ethnicity. Furthermore, to examine whether the SES of applicants' names confounds the effect of ethnicity, half of the majority names were randomly assigned from a pool of names associated with low SES. The pool of jobs applied for was diverse, representing various skill sets and spanning both the public and the private sector. The applications contained relevant experience and education as well as markers signalling that the applicants were competent, tolerant, and likeable people. In other words, the applicants were highly qualified for the jobs they applied for, which, according to previous empirical studies, is likely to minimize the employers' incentive to discriminate (Agerström et al. 2012; Kaas and Manger 2012; Birkelund, Heggebø, and Rogstad 2017).

The article reports three main findings. First, similar to other studies, we find evidence of considerable discrimination in the hiring process with a callback ratio of 1.52 for job interviews between applicants with traditional Danish-sounding names and Middle Eastern-sounding names. This means that applicants with Middle Eastern-sounding names on average have to apply for 52 per

cent more jobs to receive the same number of callbacks as applicants with Danish-sounding names. Second, the results show that discrimination varies substantially by gender, with a significantly larger ethnic difference among male applicants. The interaction between ethnicity and gender exists across sector and in gender-balanced occupations as well as in occupations that are highly gender-dominated. Third, we find no evidence that the SES of the majority applicants' names moderates the effect of the ethnic treatment, which supports the notion that the differences between treatment groups are caused by the ethnic trait and are not associated with SES.

The Danish Context

The motivation for examining discrimination against workers with Middle Eastern-sounding names in Denmark is twofold. First, there has been a rapid increase in the number of non-Western immigrants and descendants in Denmark in the past 30 years, most of whom are of Middle Eastern origin, with the largest groups being from Turkey, Lebanon, Pakistan, and Iraq. The immigration from non-Western countries was originally intended to supply the booming labour market in the late 1960s, but since the mid-1970s, most of the migration has consisted of asylum seekers and family reunifications. Today, non-Western immigrants and descendants are a significantly larger group than immigrants and descendants from Western countries and comprise approximately 8.5 per cent of the total Danish population (Statistics Denmark, 2017). Second, immigrants and descendants of Middle Eastern origin have been and still is a very salient group in the persistent political debate over immigration and integration in Denmark (Simonsen 2017). Discussions about the economic and cultural integration of immigrants of Middle Eastern origin revolve around the comparatively high unemployment rates and the fact that minorities of Middle Eastern origin fare worse on a number of socio-economic indicators such as education, income levels, and crime rates (Statistics Denmark, 2017). This is reflected in attitudes among native Danish citizens who generally recognize immigration and integration as vital political topics. Anti-immigrant attitudes in Denmark are equivalent to most other European countries (Dinesen and Sønderskov 2015), which also manifests in support for parties running on an immigration sceptical platform (Rydgren 2008;

Mudde 2013). Finally, while research on actual discrimination in Denmark is limited, studies on perceived discrimination show that large shares of non-western immigrants have experienced labour market discrimination (Jensen et al. 2012) and that employers perceive language and cultural issues as barriers when hiring immigrant-origin minorities (Slot 2008).

Existing evidence and theoretical background

A wealth of research has examined attitudes towards ethnic minorities and self-reported experiences of discrimination, but since discrimination is a sensitive topic and events can be misjudged or overlooked, it remains unclear to what extent experiences of discrimination correspond to a reliable representation of reality (Pager and Shepherd 2008). However, correspondence experiments makes seemingly ‘immutable characteristics’ manipulable by exposing units to signals of the given characteristic (Sen and Wasow 2016). The basic idea is to hold constant anything but the group characteristic being examined. By exposing employers to randomly assigned traits associated with a given group—e.g. gender, race, or ethnic categories—any difference in outcomes can be ascribed to the treatment. Conducting the experiments in the field is key to measuring actual behaviour when studying a highly sensitive topic such as discrimination.

A body of research covering a large number of countries has accumulated evidence that corroborates the existence of discrimination of outgroups in labour markets.¹ In a review of correspondence experiments conducted in the period 1990–2015, Zschirnt and Ruedin (2016) conclude that experimental research consistently finds proof of ethnic or racial discrimination in the hiring process across OECD countries. It should be kept in mind that results from different correspondence experiments cannot be compared directly, since variations in experimental designs and the demand for labour in local contexts vary across studies. Nevertheless, when averaging across studies, minorities have to apply for 50 per cent more jobs to receive the same number of job interviews as the majority group (Zschirnt and Ruedin, 2016). In summary, these differences are consistent and substantial

¹ The method has also been applied in a number of domains outside the labour market, measuring differential treatment in housing markets (Fang, Guess and Humphreys, forthcoming), the market place (Ayres and Siegelman 1995), the sharing economy (Edelman, Luca and Svirsky, 2017), or state legislators’ responsiveness to requests from voters (Butler and Broockman 2011).

across a large number of countries.

Theories of discrimination

Following the empirical evidence of differential treatment, the question of *why* ethnic or racial discrimination occurs is obviously of immense interest. Two theories of discrimination dominate the literature. In the *taste-based discrimination* model introduced by Becker (1958), discrimination is seen as the result of an irrational distaste towards certain groups. In other words, because of prejudiced employers, co-workers, or customers, there is a disamenity value to employing minority workers, resulting in preferential hiring and wage differentials (Guryan and Charles 2013). An alternative explanation for discrimination is found in the so-called *statistical discrimination* models (Phelps 1972). The foundation of these models is that employers have limited information about applicants' productivity, giving them an incentive to utilize their knowledge on the average productivity of the applicants' group in the evaluation of individual applicants. Thus, if ethnicity correlates with undesired traits, discrimination based on ethnicity becomes an optimization strategy (Arrow et al. 1973).

While the theoretical premises of the two models of discrimination are fundamentally different, it has proven difficult to empirically differentiate between them (Bertrand and Mullainathan, 2004; Dancygier and Laitin, 2014). First, and most fundamental, it is difficult to elicit distinct observable implications between the two theories, and therefore to distinguish between them empirically.² Second, different types of discrimination might interact over time. Disadvantages initially caused by taste-based discrimination can eventually initiate real group differences in education or labour market outcomes, creating a basis for statistical discrimination. Even if one type of discrimination is successfully identified at one point in time, it can be a complex result of preceding processes (Dancygier and Laitin, 2014). With these limitations in mind, well-designed correspondence experiments enable researchers to identify which components trigger discrimination (Sen and Wasow,

² One way of assessing the implications of theories of discrimination has been to examine heterogeneous treatment effects—e.g. varying effect sizes across firm size or customer contact. While such treatment-by-covariates effects can be interesting, they are not solid answers to causal questions. Jobs with certain characteristics might differ systematically on a number of unobserved variables that alter the explanation.

2016). When the particular ways in which group characteristics, alone and in combination, result in disparities is understood, discrimination is much more likely to be recognized and addressed.

Intersections between ethnicity and gender in the hiring process

While the combination of gender and ethnicity in labour market discrimination has been under-researched in the field experimental literature, the broader literature on intersections between gender and ethnicity is vast. Two of the most prominent notions about gendered ethnic discrimination outline very different empirical implications; first, from the perspective of intersectionality, it is argued that minority women will suffer the largest disadvantage, since they occupy the lowest position in both social categories, being female and belonging to an immigrant-origin minority group (Ransford, 1980; Harnois, 2015). This dual oppressive system—whether it is the notion of additive jeopardies (Beal, 1970) or multiplicative jeopardies (King, 1988)—can be translated into a ‘supplementary discrimination hypothesis’ that expects a larger ethnic gap among female applicants.

An alternative prediction is found in the social psychology literature. Social dominance theory, a general model of hierarchically structured relationships among social groups, argues that ethnic conflict is primarily executed by and targeted against males (Sidanius and Pratto 2001). It is claimed that minority males are perceived as a greater threat and therefore are the primary target of discrimination, while outgroup females are less susceptible to discrimination. This is also known as the ‘outgroup-male-target hypothesis’ (Navarrete et al., 2010). Arguably, both notions can be understood along the lines of taste-based discrimination, with a focus on factors unrelated to the productivity-related characteristics. However, interactions between ethnicity and gender can be a result of statistical discrimination too. For example, minority males in general have lower educational credentials than minority females, and they are over-represented in the criminal justice system, which can be expected to translate into different levels of discrimination.

Although the majority of European correspondence experiments rely on male applicants, there is good reason to consider the importance of gender in studies on ethnic discrimination. Evidence from labour market field experiments that manipulate applicants’ gender is scarce and findings

are mixed. While some research supports the outgroup-male-target hypothesis (Andriessen et al., 2012; Arai, Bursell and Nekby, 2016; Liebkind, Larja and Brylka, 2016; Midtbøen, 2016), the findings are not consistent. Some correspondence experiments find little or no variation in ethnic discrimination across gender (Blommaert, Coenders and van Tubergen, 2014; Bursell, 2014; Derous, Ryan and Nguyen, 2012). Moreover, recent studies emphasize how intersections between ethnicity and gender differ substantially across different occupations. In a study conducted in the Swedish labour market, Bursell (2014) finds a larger ethnic difference among males in male-dominated occupations, and in the Norwegian labour market, Midtbøen (2016) finds that the effect of the ethnic trait is larger among male applicants, but not in gender-integrated occupations in the private sector. This is especially important, since most of the aforementioned studies adjust the research design according to patterns in gender stratification such that job openings within occupations that are very male-dominated only received applications by male applicants and vice versa.³ In this study, we randomly assigned pairs of applicants with the same gender to each job, which allows us to compare callback rates across ethnicity and gender without any adjustments to the types of job applied for. In addition, we can assess effect heterogeneity by breaking down the results across occupations according to gender composition.

Using names as proxies for ethnicity and potential confounding

Correspondence studies examining ethnic discrimination rely on the assumption that differences in callbacks are exclusively due to the signal that the name provides about ethnicity. However, perceptions about names might be influenced not only by the population racial/ethnic composition of a name but also its population SES (Gaddis 2017). This becomes an issue in correspondence studies where distinct minority names might be perceived as low-status compared to the distinct majority names. Hence, it is possible that studies relying on racial or ethnic distinct names are picking up a confounding relationship between ethnicity/race and SES (Fryer and Levitt, 2004).

³ One argument for doing so is to avoid evoking suspicion among employers in gender-stratified occupations where two applications from similar qualified applicants of the under-represented gender will seem odd (Arai, Bursell and Nekby, 2016). Another argument relates to the real-life consequences of discrimination: if very few female candidates work in construction, the need to examine and address discrimination against female candidates is arguably smaller.

Thus far, this notion is largely theoretical and the evidence supporting it is limited and ambiguous.⁴ In the American context, two recent correspondence experiments in the labour market did not replicate the previous results of racial disparities (Darolia et al. 2016; Deming et al. 2016). One possible reason for this is that the names used to signal race had different connotations of SES than previous correspondence experiments. On the other hand, the divergence from previous research could also be due to employers not being able to precisely identify applicants' race, which attenuates the effect.⁵

In a European context, Jackson (2009) conducted a field experiment in the United Kingdom to examine if different traits signifying social class had an effect on callbacks from employers. Overall, the combination of different high-status characteristics only resulted in small advantages, but the largest observed (positive) effect of an individual treatment, although only borderline significant, was attributed to holding an elite name. Hence, there is good reason to examine potential effects of SES related to names to obtain a valid measure of ethnic discrimination and provide new information for the discussion of the importance of SES.

Since Middle Eastern-minorities in Denmark fare worse on a number of socio-economic indicators compared to native Danes, it is plausible that employers perceive them as a low-status group. One way to address this in the research design is to match majority and minority applicants on social class (Gaddis and Ghoshal 2015). However, since there is no reason to believe that employers are able to distinguish between high- and low-SES minority names, it is only possible to manipulate the SES component among majority applicants.⁶ The section on the study's research design outlines the strategy for choosing the specific names.

⁴ Butler and Homola (2017) perform an ex post analysis of an audit study on political responsiveness using public records to assess the importance of SES and political resources reflected in names. They find no evidence that these signals predict the probability of legislators' likelihood of responding.

⁵ To circumvent the risk of ascertaining SES of racially distinct first names, Darolia et al. (2016) only signify race through distinct surnames, which might not be a clear signal of race.

⁶ We initially assumed that the large majority of employers would belong to the majority group. From the names of employers or HR managers that we contacted, we only identified one with a minority name.

Hypotheses

Based on the discussion in the previous section, we examine three hypotheses that were all pre-registered at the EGAP.org database.⁷ In answering these hypotheses, we rely on the framework outlined by Sen and Wasow (2016), conceptualizing ethnicity as a composite variable rather than a single uniform entity. In this framework, ethnicity is a fusion of several factors, such as region of ancestry, religion, or SES, which might trigger discrimination in different ways. By exposing employers to different manipulations of randomly assigned characteristics, we can disentangle how different aspects alone and in combination affect behaviour. This is a useful methodological starting point that enables a deeper understanding of which and when social groups are subject to differential treatment.

First, we are interested in the overall difference, *ceteris paribus*, between the two ethnic groups. Based on the consistent findings of ethnic discrimination in previous European correspondence experiments in mind, we hypothesize that there will be an overall significant difference between the majority group and the immigrant-origin minority group:

H1: *Applicants with Middle Eastern-sounding names are less likely to receive a callback than applicants with traditional Danish-sounding names.*

Since perceptions of males and females within the same ethnic group might differ, it is essential to include both groups in the treatment to get a general measure of ethnic discrimination. Following the theoretical and empirical work outlined in the previous section, we test the gendered nature of ethnic discrimination in a second hypothesis:

H2: *The difference in callbacks between majority and minority applicants is larger among male applicants than among female applicants.*

Finally, we examine if the SES of the names used to signal ethnicity might be a confounding variable

⁷ See details at Supplementary Appendix G.

by manipulating the SES of the majority applicants' names.

H3: *The difference in callbacks between majority and minority applicants is larger when the majority applicant holds a popular name than when the majority applicant holds a low-SES name.*

Experimental design and implementation

The experimental design in correspondence studies, especially the quality of the applications and the jobs applied for, can influence results substantially (Neumark, 2012). In this study, we aimed for a design that minimizes employers' incentive to discriminate against the minority applicants. These design features and the implementation of the experiment are described in detail in the following sections.

Treatments and randomization details

We randomized the assignment of three different treatments—*ethnicity*, *gender*, and *SES*—using names as proxies. The applicants' gender was also explicitly stated in the CVs to avoid potential misconceptions of the treatment. Each job received two applications, one with a Danish-sounding name and one with a Middle Eastern-sounding name. Gender was randomly assigned pairwise, so the applicants for any given job were either two males or two females. In other words, we randomized ethnicity *within* and gender *across* the job ads (Figure 1). Finally, to study if SES affects chances of a callback, the traditional Danish-sounding names were randomly assigned from two pools: either the most popular Danish names or Danish low-status names.

Presumably, it is difficult for most employers to differentiate the SES of various Middle Eastern-sounding names, which is why we only manipulated the SES of the pool of Danish-sounding names. We can test the importance of SES by comparing, on the one hand, the difference in callbacks between minority applicants and applicants with the most popular Danish-sounding names, and, on the other hand, the difference in callbacks between minority applicants and applicants with low-status Danish-sounding names.

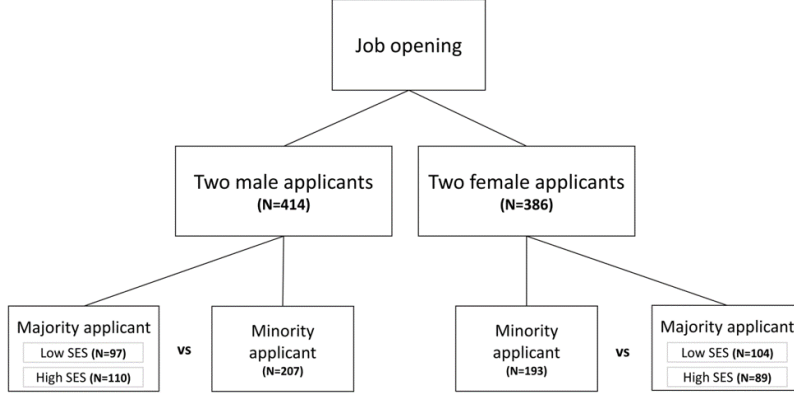
We used a large number of different names (so-called stimuli sampling) to ensure that effects are caused by the categories of interest and not the unique characteristics of a specific name (Wells and Windschitl, 1999). The names can be divided into three groups: a pool of the most common Danish-sounding male and female first names, a pool of the most common male and female first names used by Middle Eastern minorities in Denmark (Statistics Denmark, 2015), and a pool of Danish male and female low-status names. Furthermore, two pools of surnames with the most common traditional Danish-sounding (Middle Eastern-sounding) surnames were randomly paired with the pool of Danish-sounding (Middle Eastern-sounding) first names.⁸

It is essential that the names serve as distinct signals of ethnicity and gender (Riach and Rich 2002). The immigrant-origin minority names were identified from a dictionary of Middle Eastern first names that are in frequent use in Denmark (Meldgaard, 2005) and are among the 20 most popular Middle Eastern-sounding names in Denmark. Furthermore, since all candidates have a distinct Middle Eastern-sounding first name *and* a Middle Eastern-sounding surname, we effectively rule out any confusion about the precision of the ethnic signal. It might be expected that some minority names evoke stronger Muslim connotations (e.g. Mohammed) that result in larger ethnic penalties, but callbacks are evenly distributed across names (Supplementary Appendix J).

To identify the pool of low-SES names, we used a data set on the social characteristics of Danish citizens. We constructed an index of SES by using data on average income, crime rates, and unemployment for Danish citizens with a given name.⁹ Furthermore, we excluded first names with an average age below 20 years or above 40 years. Hence, we identified the low-SES names from the bottom decile of the index and chose American-inspired names that fit a common Danish stereotype of being low-status (see details in Supplementary Appendix A)

⁸ We used data on the most common Danish names from Statistics Denmark (2015). See a list of names in Supplementary Appendix A.

⁹ The data set included register data on more than 3.8 million Danes.

Figure 1. Assignment to treatment and control groups

Constructing applications and CVs

The applications were designed to meet three criteria: *(i)* they should be perceived as real applications, *(ii)* each application in a pair should be different to avoid arousing suspicion, and *(iii)* the applicants should be equally qualified (Midtbøen and Rogstad, 2012). We adhered to the standard procedure in the Danish labour market by submitting a one-page application letter and a CV with the inclusion of a phone number and email. We constructed the generic applications building on four paragraphs: introduction, motivation, experience, and personal interests. For each paragraph, we constructed two slightly different texts (A and B). By utilizing all possible combinations of the A and B paragraphs, we got 16 unique pairs of applications where each pair consisted of each other's opposite (e.g. one possible pair is A-A-B-B and B-B-A-A). We randomized the assignment of CVs to each application and randomized which application to send first (the elapsed time between the two applications was 2 days). There were no differences in callbacks for job interviews across the different application pairs and CVs, which supports the notion that employers perceived them as equally qualified (Supplementary Appendix B).

The applications were largely generic but held a few empty spaces that were completed in each specific case to match an application with a particular job (See Table 1). For example, the applicant states: 'I enjoy working [...]', and we added 'independently' or 'in a team' depending on

the given job. If the job ad explicitly requested specific requirements, these were incorporated into the CVs (software skills, a driver's licence, etc.). The modification of the applications for each job secured a sufficient chance of receiving a callback. It is impossible to know the relative quality of the applications, since it depends on other applicants applying; however, the absolute quality can be adjusted (Neumark, 2012). While underqualified applications will receive few or no answers, over-qualified applications will rule out statistical discrimination by receiving a callback too often.

Broadly speaking, qualifications can be divided into cultural characteristics and professional characteristics. Our applicants have strong professional characteristics, possessing the necessary educational background and relevant experience from previous jobs.¹⁰ The cultural characteristics are signalled in the perfectly written Danish and the fact that applicants love to cook with their boyfriend/girlfriend, do triathlons, or jog with good friends. These markers of a Danish middle-class lifestyle were included to reduce the perceived cultural differences between the applicants, hence minimizing the incentive to discriminate against the minority applicant. Finally, in the CV it was stated that the applicant was born in 1984 and the gender of the applicant was indicated explicitly. In summary, the applicants were highly qualified for most of the jobs applied for, which is also indicated by the high callback rates. Following previous studies, the high standard of the applications can be expected to reduce the ethnic discrimination (Kaas and Manger, 2012; Birkelund et al., 2014). As such, the study constitutes a least-likely case for observing differential treatment.

¹⁰ The applicants' CVs mentioned real educational institutions in the section on educational background and real workplaces in the section on experience to maximize realism. We did not receive any comments indicating that employers had been in contact with institutions or workplaces.

Table 1. Assignment to treatment and control groups

	A	B
Introduction	I believe the job as [] represents an excellent opportunity for me. I am highly motivated and have experience working as []. I am currently employed at [], but I recently moved to [] which is why I am looking for a job closer to my new home.	I would like to put forth my candidature for the position as []. I am very motivated and believe I have the experience to meet the requirements. I am currently employed as [] but after [] years I now feel it is time to look for new challenges.
Motivation	I enjoy working [<i>in a team/on my own</i>] and I will thrive in a position as []. I am familiar with [] and know you are among the [<i>most ambitious/most professional/best</i>] workplaces. I am very passionate about my work and would love to be part of a team with a professional and positive attitude.	Working [<i>independently/ with good colleagues</i>] motivates me, and I see the job at [] as a great match. [] is without doubt a great place to work and I would love to be part of that. I am devoted to my career and it is essential to me that I carry out a professional job. I have managed to do so in my previous jobs and hope to continue at [].
Experience	I am a very ambitious [] and I work hard to achieve good results. In my previous jobs, I have always been given responsibility and I have been an esteemed employee. I hope that I can achieve the same in the job at []. From [<i>year</i>] to [<i>year</i>] I worked in [] with []. The last [] years I have worked in [] why I bring experience to the job.	I have always worked hard and with great dedication. As [] I have been a trusted employee with a lot of responsibility. In my jobs at [] and [] where I worked for [] and [] years, respectively, we managed to [<i>increase sale/provide a great service/do a professional job each day</i>]. I hope to continue in the job at [].
Personal information	I am a very positive and social person who believes good relationships in the workplace is important. In my spare time, I do a lot of sports and run several times a week together with good friends. Besides that, I love to travel and scuba dive with my boy/ <i>girl</i> friend. I am very excited to learn more about this opportunity and share how I will be a great fit for [] Sincerely, [<i>Name</i>]	In my spare time, I love to do sports and I do triathlon in the summer season. I also spend a lot of time cooking with my boy/ <i>girl</i> friend. I am an optimistic person with a good sense of humour. Having a good relationship with my colleagues is important to me. I would appreciate the opportunity to meet with you and discuss how my qualifications can benefit your organization [<i>company</i>]. Sincerely, [<i>Name</i>]

The sample

The empirical analysis builds upon experimental data collected between September 2015 and June 2016. In total, 800 applications were sent in response to 400 job openings. We sampled the jobs from the online employment portal, Jobindex.dk. Jobindex is the largest employment portal in Denmark and covers a broad array of jobs across 10 occupational categories.

To provide a comprehensive picture of potential discrimination, we applied for 23 different types of occupations within six different occupational categories: Office and administrative support, Education, Health care, Retail, Construction and extraction, and Marketing and sales (see all occupations in Supplementary Appendix D). We maximized the geographical variation and applied for jobs from all five Danish regions. In total, 278 of the ads were private sector jobs. Hence, the

sample comprises a broad geographical scope and covers a relatively large segment of the Danish labour market.

We excluded academic jobs as well as highly technical jobs from the sample, since these would require detailed knowledge of essential skills, specific applications, recommendations, and lengthy CVs that would exceed the generic applications used in this experiment. The sample includes occupations that are dominated by either women or men and occupations where the labour force is mixed. We coded all jobs according to the gender composition in the specific occupation. Using a threshold of 20 per cent, we identified 112 jobs as female-dominated and 90 jobs as male-dominated, while the remaining 198 jobs had at least 20 per cent male and female workers (Supplementary Appendix D). To minimize the risk of receiving an answer before sending the second application, we only applied for jobs where the expiration date from the job advertisement was 2 weeks or more. Some job ads demanded a picture of the candidate, in which case we did not apply.

Randomization and handling of answers

For each job ad, we initially registered background information (sector, number of employees, language requirements, and contact information) and adapted the application and CV templates to the specific job before both applicants were finally randomly assigned a gender and each application was randomly assigned an ethnic affiliation. By finishing both applications first and randomly assigning names afterwards, we avoided the risk of unintentionally biasing the quality of the applications. If employers contacted one or both of the applicants, the job interview offer was politely declined. We define a callback as a personalized contact in the form of a message on the answering machine or an email from the potential employer with an invitation for a job interview. In a few cases, employers contacted applicants with additional questions or they asked for proof of education, which did not qualify as a callback. All communications with employers were archived on either email or answering machines. In 178 of 222 callbacks, we received an email.

Ethical considerations

There is a large body of literature on the ethical considerations involved when conducting correspondence experiments, including advice on how to minimize harm to subjects involved (see Riach and Rich (2004) and Zschirnt (2016) for an extensive discussion of ethics in correspondence experiments). Central arguments for the legitimacy of carrying out correspondence experiments include (i) the question of discrimination is of high societal importance, (ii) that there is no other way to credibly retrieve this vital information, and (iii) if the research is prepared and carried out carefully, there is only a very limited detrimental effect on the employers tested (Riach and Rich 2004; Zschirnt 2016).

On the last point, we conducted a pilot study to confirm that the experimental set-up and the logistic of sending and handling the callbacks operated efficiently. Furthermore, to ensure sufficient statistical power without having to contact an excessively large number of employers, we conducted a power analysis before the collection of data. Finally, we carefully considered how to minimize any inconveniences to the employers that were contacted as a part of the experiment. The main cost to employers is time, so we answered all requests as quickly as possible, explaining that the applicant had just found another job. We did not debrief employers to avoid the risk of making them doubt the credibility of future (minority) applicants. Furthermore, we analysed data on an aggregated form to ensure the anonymity of all individual employers to avoid associations between specific companies and this study. Finally, the data is kept in encrypted files on a protected server.

Balance Check

We performed a balance check to judge whether the random assignment procedure seems to be reliable (Gerber and Green 2012). The main treatment, ethnicity, is necessarily balanced across covariates, since all job ads received an application from both a majority and a minority applicant. However, this is not the case for gender and SES, and therefore, we tested whether observed covariate imbalances are larger than expected from chance alone. To examine this, we regressed the

treatment indicators (gender and SES, respectively) on the available covariates¹¹ and calculated the heteroscedasticity-robust Wald statistic for the hypothesis that all the coefficients on the covariates are 0 (Lin, Green and Coppock, 2016). To calculate the related P-values, we used randomization inference to create a distribution of Wald statistics under the null hypothesis of no systematic imbalance. The results show no reason to reject the null hypothesis that the pre-treatment covariates are not systematically related to the treatment (P -values: 0.43 and 0.92), and we therefore consider the assignment of treatments to be balanced (see Supplementary Appendix C for further details).

Results and interpretation

To recap, this experiment answers three main questions. First, are equally qualified applicants with either a Danish-sounding name or a middle Eastern-sounding name treated differently by employers? Second, is the effect of having a minority name moderated by applicants' gender? Finally, is ethnic discrimination exclusively caused by the ethnic trait or does SES confound the effect? All P-values and standard errors reported in the analysis are obtained from randomization inference with 100,000 iterations.¹²

Main Results

In total, 800 applications were sent to 400 jobs and at least one applicant received a callback with an invitation for a job interview in 39.5 per cent of these. Since each job opening received two equally qualified applications, we can observe two potential callbacks for each workplace. As is evident from Table 2, the callbacks were not equally distributed, with a substantial gap between minority and majority applicants. The majority applicants received a callback rate of 33.5 per cent, which is a substantially higher share than the minority applicants, who received a callback on 22 per cent of the applications. The difference corresponds to a ratio of 1.52, implying that minority

¹¹ Covariates include Size (number of employees); Sector (public/private); Education (if education after high school was needed); Language required; and Customer contact.

¹² By reproducing the randomization procedure a large number of times, the distribution of the test statistic under the sharp null hypothesis can be approximated with a high degree of precision (Gerber and Green, 2012).

applicants need to send 52 per cent more applications to receive the same number of invitations as applicants with traditional Danish names. In the literature there are different ways to report outcomes from correspondence experiments. In this study, we first and foremost pay attention to the relative callback ratio and the difference in means (DIM), but Table 2 also reports the level of net discrimination, which is a common measure of discrimination in the literature. For all applicants, the net discrimination rate is 0.29, while it is 0.41 and 0.15 for male and female candidates, respectively.¹³

Table 2. Callbacks across ethnicity and gender

	Males	Females	Total
Callback majority	36.2	30.6	33.5
Callback minority	19.3	24.9	22
Ratio	1.88	1.23	1.52
DIM	16.9*** (3.66)	5.7* (3.15)	11.5*** (2.41)
Neither invited	121	121	242
Both invited	29	35	64
Only majority invited	46	24	70
Only minority invited	11	13	24
Net discrimination	0.41	0.15	0.29
N	207	193	400

Note: Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 2 also reports the DIM estimates and the related standard errors obtained from randomization inference with 100,000 iterations under the sharp null hypothesis, assuming no effect of ethnicity for all applicants (Supplementary Appendix F). It is extremely unlikely ($P < 0.001$) that

¹³ The measure of net discrimination treats cases with no callbacks as non-observations and is obtained by dividing the difference between observations where only the majority was invited and observations where only the minority was invited with the number of observations where at least one candidate was invited.

the overall ethnic difference in means of 11.5 percentage points would have occurred by chance. If we break down the results into occupational categories, we see that although the relative difference varies, the majority applicant is preferred over the minority applicant in all six occupational categories (see Supplementary Appendix D for details). Overall, the results suggest that employers across occupations use ethnicity as an important decision rule when evaluating applications, and hence, that applicants with a Middle Eastern background are subject to discrimination

Callbacks across gender and ethnicity

From the results reported in Table 2, it is noticeable that the ethnic difference in callbacks seems to be gender-reliant. The results are visualized in Figure 2A, showing the difference in callbacks between majority and minority applicants for female applicants, male applicants, and all applicants, respectively.

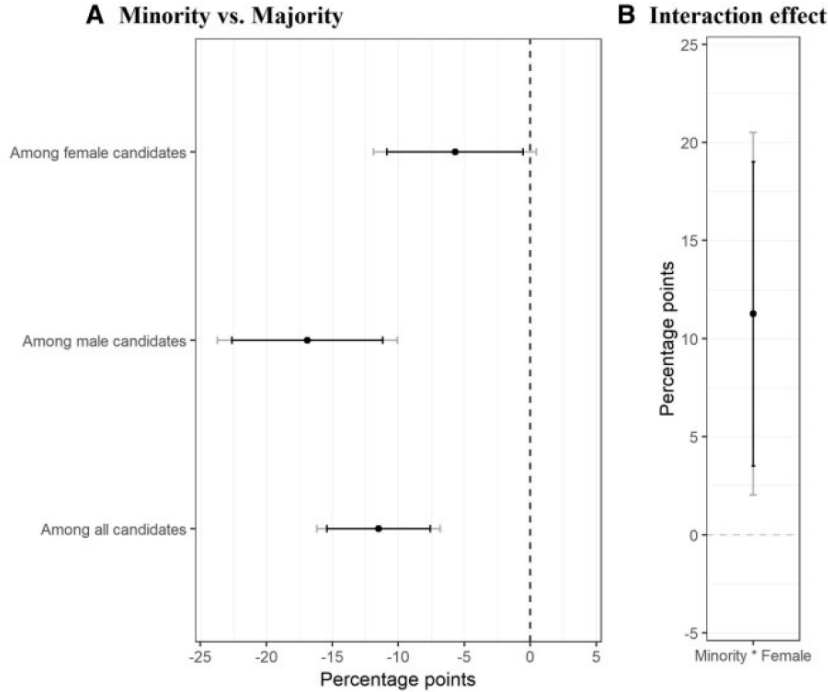
To test if the interaction between ethnicity and gender is significant, we regress a callback dummy on ethnicity and gender of the applicant as well as the interaction between the two.¹⁴ H2 implies that the interaction term between immigrant-origin minority and female should be positive and significant. As is evident from Figure 2B the estimate of the interaction effect is noisy, but the effect is substantial (11.2 percentage points) and statistically significant ($P=0.016$).

While there is a large penalty for belonging to the immigrant-origin minority group and a small (insignificant) penalty for being female within the majority group, these differences are not additive. Instead, the interaction term denotes that minority females receive a substantively higher callback rate than we would expect if the ethnic and gender differences were purely additive.

As pointed out previously, heterogeneous effects across occupations can be decisive for the overall effects in correspondence experiments (Bursell, 2014; Midtbøen, 2016). Could the interaction effect between ethnicity and gender to some extent be an artefact of the sample's composition of

¹⁴ The fact that ethnicity is randomly assigned within occupations and gender between occupations implies that the effect of ethnicity is measured with more precision than the effect of gender. We account for this by using randomization inference with the same randomization procedure. Gender is clustered on the job level and ethnicity is block randomized on the job level. We impute constant additive effects and run 100,000 iterations. Alternative specifications using ordinary least squares regression with clustered standard errors generate similar results (Supplementary Appendix F1).

Figure 2. (A) Difference in callbacks between majority and minority candidates across gender. (B) The interaction between gender and ethnicity.



Note: CI are shown at the 95 and 90 per cent level.

occupations? We explore this notion in two steps. First, we re-weight callback rates by occupational categories and examine the results given the sample consisting of equally sized occupational categories. As is evident from Table 3, although the relative differences are slightly smaller in the re-weighted sample, the same overall pattern of gendered differences in ethnic discrimination is consistent. For female applicants, the re-weighted ratio is 1.18 compared to 1.22 in the original sample, while for male applicants the callback ratio alters from 1.88 in the original sample to 1.76 in the re-weighted sample. The total ratio between majority and minority applicants is reduced from 1.52 to 1.44.

Secondly, we explore results in subsets of the sample based on the gender stratification in the labour market. We identify 112 jobs in female-dominated occupations (share of males < 20

Table 3. Callback rates re-weighted by occupational category

	Female	Male	Total
Majority	31.5	38.1	34.9
Minority	26.7	21.6	24.1
Ratio	1.18	1.76	1.44

pct.) and 90 jobs in male-dominated occupations (share of females < 20 pct.), while the rest is denoted as gender-balanced occupations. Female-dominated occupations in the sample include jobs as nurses, pedagogues, social and healthcare assistants, social workers, receptionists and secretaries, and cleaning workers, while male-dominated occupations include building and construction workers, mechanics, warehouse workers, IT supporters, and drivers (Supplementary Appendix D).

As reported in Table 4, the overall gendered difference in ethnic discrimination is consistent in all three subsets of the data. In each subset, minority males received fewer callbacks than both majority males and majority females. Hence, the ratios are substantially smaller among female candidates vis-à-vis male candidates. The relative difference between majority and minority applicants is most substantial in the male-dominated jobs, where minority males received a callback rate of 9.6 pct. compared to 36.5 pct. of the majority males, implying a relative difference of 3.8. The smallest ratio of 1.19 is observed among female applicants in female dominated occupations where the ratio is 1.73 among male applicants.

Table 4. Callbacks across gender-segregated occupations

	Gender-balanced occupations		Male-dominated occupations		Female-dominated occupations	
	Female	Male	Female	Male	Female	Male
Majority	29	37.8	28.9	36.5	34.5	33.3
Minority	24	24.4	21	9.6	29.1	19.2
Ratio	1.2	1.54	1.38	3.8	1.19	1.73
N	100	98	38	52	55	57

In summary, the exploratory analysis shows that the effects do not seem to be an artefact of the composition of occupational categories in the sample. Moreover, the interaction between gender and ethnicity is present in both gender-balanced, male-dominated and female-dominated occupations. However, the relative difference among majority and minority applicants is substantially larger in male-dominated jobs, suggesting that males both execute and become the target of ethnic discrimination more often than females.

Does SES confound the use of names as a signal for ethnicity?

To examine if the ethnic traits are confounded by notions of SES, we randomly assigned the majority names from two groups: the first group consisted of the most popular Danish names, while the other group consisted of names from the bottom percentile of the SES index. On this basis, we test if the difference between immigrant-origin minority and majority candidates is confounded by the status of the majority applicants.

We seek to estimate the differences in treatment effects of ethnicity conditional on the information about SES related to the majority applicants. Hence, the quantity of interest in this section is a difference in differences.¹⁵ This estimand captures the degree to which effects of SES are consequential for the relative difference in callbacks. If having a low-SES name reduces the chances of receiving a callback, the estimand should be positive.

Table 5 shows callback rates for competing pairs of majority and minority applicants. The immediate comparison of callbacks to majority applicants indicates that having a low-SES name reduces the chances of receiving a callback by roughly 4 percentage points. However, the relative difference in callbacks between majority and minority applicants is actually slightly larger when the majority candidate has a low-SES name (12.4 percentage points) compared to jobs where the majority candidate has a popular name (10.6 percentage points).

¹⁵ This estimand is represented by the following expression:

$$(E[\text{Callback} \mid \text{Majority} \mid \text{popular}] - E[\text{Callback} \mid \text{Minority competing against Majority popular}]) \\ - (E[\text{Callback} \mid \text{Majority lowSES}] - E[\text{Callback} \mid \text{Minority competing against Majority lowSES}])$$

Table 5. Callbacks conditioned on SES of the majority applicant

	Competing pairs when majority candidate has a popular name	Competing pairs when majority candidate has a low-SES name
Majority	35.7	31.3
Minority	25.1	18.9
Difference in means	10.6*** (3.2)	12.4*** (3.3)
N	198	202
Difference in differences		−1.8 (4.8)

Note: *p<0.1; **p<0.05; ***p<0.01.

Hence, the difference in differences is negative, which suggests that that ethnic discrimination is not altered by the status of the majority candidate. While the estimate is imprecise, we can reject the third hypothesis that SES related to majority applicants' names is a significant factor.

One concern is that our sample size may be too small to detect effects of SES, but the fact that the difference in differences is actually negative supports the conclusion that providing information on SES among majority candidates does not substantially affect levels of ethnic discrimination. We also explore the intersections between gender, SES, and ethnicity but find no major differences from the overall pattern (see Supplementary Appendix H).

In conclusion, the excludability assumption seems to hold: the differences between ethnic majority and immigrant-origin minority candidates are not affected by the SES related to the majority applicants' names. This bolsters the validity of the use of distinctive names to signal ethnicity.

Conclusion and discussion

This article sheds new light on the disparate treatment of immigrant-origin minorities in the labour market, a topic that has received substantial interest in the scholarly community as well as having

been the subject of intense public debate. The results show that when equally qualified applicants apply for a job, an applicant with a Middle Eastern-sounding name is significantly less likely to receive a callback compared to an applicant with a Danish-sounding name. The overall result is strikingly similar to the findings of previous correspondence studies conducted in a number of European countries (Zschirnt and Ruedin, 2016). The difference is particularly notable in light of the fact that the fictitious applicants used in this experiment were highly qualified for the jobs applied for, which should have minimized the incentive to discriminate.

The findings demonstrate that discrimination varied substantially by gender, which serves as a call to future correspondence experiments to manipulate gender to estimate general measures of ethnic discrimination. The results undergird the notion that male immigrant-origin minorities are particularly vulnerable to discrimination. Hence, this experiment provides empirical support for the outgroup-male-target hypothesis. Contrary to previous studies on gendered ethnic discrimination (e.g. Andriessen et al. (2012) and Midtbøen (2016)), the results indicate that such discrimination is not limited to specific occupational categories but is instead consistent across a broad spectrum of the labour market. Additional research is required to explore the conditions under which ethnic discrimination is gender-reliant.

Furthermore, this article examined whether the SES of names influences the chances of receiving a callback. By actively choosing control names, we were able to examine if the ethnic gap decreased when the majority name belonged to a group of low-SES names. The results do not suggest that the majority applicants' names confounded the effect of ethnicity, which bolsters the interpretation that the gap in callbacks can be attributed to ethnicity and not characteristics related to the use of distinct names.

Despite its innovations, the present study also has a number of limitations. First, this experiment proves discrimination against candidates with a given set of credentials in a non-representative sample of the Danish labour market. While we did apply for a broad array of jobs, it is impossible to draw a representative sample of job openings from the ever-changing job market. Hence, applicants applying for other types of jobs—and with different educational levels, experience or personal characteristics—might face different outcomes than found in the present study. It should also be noted

that while we have studied discrimination in the first stage of the hiring process, discrimination could occur at the job interview or within a workplace in the wage-setting or promotion process (Pager and Western 2012).

Second, despite the large difference in invitations for job interviews, it remains uncertain how discrimination affects the employment rate of the large group of Middle Eastern immigrants and descendants in Denmark more generally. Because the availability of jobs and the strategy of the individual applicant play a decisive role in getting a job, discrimination on average does not necessarily translate into gaps in employment rates (Heckman 1998). Minorities might have prior knowledge about non-discriminating workplaces or find jobs outside of the formal hiring processes (Demireva, 2008), and thereby avoid discriminating employers. Nevertheless, the gap between majority and minority candidates found in this experiment is substantial and occurred across different occupational categories, from the largest Danish online job portal, and it therefore seems highly plausible that ethnic discrimination translates into overall employment disadvantages. Furthermore, the findings raise concerns about the presence of a meritocratic principle and suggest widespread violations of the law of equal treatment in the labour market.

Third, a concern in correspondence experiments is that names are imprecise proxies of the groups they represent. In this case, the names used to signify gender and ethnic differences are distinct, and there is little reason to doubt that employers understand the traits. However, it could be argued that the names used to signify SES do not match employers' perceptions of low-SES names. The names were selected based on three socio-economic factors (average crime, income, and unemployment), and even though the names match a common notion of low-status names, we cannot be completely sure that these names were perceived as such. It should be mentioned, however, that the manipulation of SES served to bolster the validity of the ethnic trait—we were not interested in studying the effects of variations in SES in itself, in which case we would have manipulated more than just the name (Jackson, 2009). In addition, it should be mentioned that while we examined the importance of having a low-SES majority name, the application templates included information that signalled a middleclass lifestyle, which might crowd out the effect of low-SES names. Hence, we cannot ignore the possibility that SES might be important if the applications

were less informative.

Finally, our findings leave open the important question of why we see these results. This experiment was primarily designed to examine if, and against whom, employers discriminated, and we cannot definitively establish the causal mechanisms underpinning the results. On the one hand, the fact that ethnic discrimination is gender-reliant supports the notion of Social Dominance theory, which can be understood as a type of taste-based discrimination mechanism. On the other hand, this finding is seemingly consistent with the expectations of statistical discrimination models as well. If minority males perform, on average, worse than minority females do on some outcome-relevant characteristics, we might expect employers to treat the two groups differently. Female descendants with a Middle Eastern background outperform male descendants in a number of statistics (education, grades, wages, etc.) which employers might utilize in their evaluation of applicants (Statistics Denmark, 2017). In addition, employers' perceptions of cultural distance might vary by the gender of the minority applicant. Cultural or value differences, for example views on gender equality, could be perceived as larger or more problematic when the minority candidate is male (Lancee et al. 2017).

While the importance of hard skills (e.g. increased experience or reference letters) has been examined in previous studies (Kaas and Manger, 2012; Arai, Bursell and Nekby, 2016), little is known about the ways in which perceptions of cultural differences moderate ethnic discrimination. This could be examined in future research by manipulating cultural information, for example by signifying support for gender equality, democratic participation, or religious affiliation. Thus, an important task in future research is to increase the understanding of why specific groups of immigrant-origin minorities are penalized. As this article demonstrates, it is possible within the framework of correspondence experiments to study how different components affect behaviour among employers. We urge researchers on ethnic discrimination to replicate and extend work in this area by disentangling the effects of other components alone and in combination to contribute towards measuring and understanding ethnic discrimination.

References

- Agerström, Jens et al. (2012). “Warm and competent Hassan= cold and incompetent Eric: A harsh equation of real-life hiring discrimination”. In: *Basic and Applied Social Psychology* 34.4, pp. 359–366.
- Andriessen, Iris et al. (2012). “Ethnic discrimination in the Dutch labor market: Its relationship with job characteristics and multiple group membership”. In: *Work and Occupations* 39.3, pp. 237–269.
- Arai, Mahmood, Moa Bursell, and Lena Nekby (2016). “The reverse gender gap in ethnic discrimination: Employer stereotypes of men and women with Arabic names”. In: *International Migration Review* 50.2, pp. 385–412.
- Arrow, Kenneth et al. (1973). “The theory of discrimination”. In: *Discrimination in labor markets* 3.10, pp. 3–33.
- Ayres, Ian and Peter Siegelman (1995). “Race and gender discrimination in bargaining for a new car”. In: *The American Economic Review*, pp. 304–321.
- Baert, Stijn et al. (2015). “Is there less discrimination in occupations where recruitment is difficult?” In: *ILR Review* 68.3, pp. 467–500.
- Bertrand, Marianne and Esther Duflo (2017). “Field experiments on discrimination”. In: *Handbook of Economic Field Experiments*. Vol. 1. Elsevier, pp. 309–393.
- Birkelund, Gunn Elisabeth, Kristian Heggebø, and Jon Rogstad (2017). “Additive or multiplicative disadvantage? The scarring effects of unemployment for ethnic minorities”. In: *European Sociological Review* 33.1, pp. 17–29.
- Bursell, Moa (2014). “The multiple burdens of foreign-named men—evidence from a field experiment on gendered ethnic hiring discrimination in Sweden”. In: *European Sociological Review* 30.3, pp. 399–409.
- Butler, Daniel M and David E Broockman (2011). “Do politicians racially discriminate against constituents? A field experiment on state legislators”. In: *American Journal of Political Science* 55.3, pp. 463–477.

- Butler, Daniel M and Margit Tavits (2017). “Does the Hijab Increase Representatives’ Perceptions of Social Distance?” In: *The Journal of Politics* 79.2, pp. 727–731.
- Carlsson, Magnus and Dan-Olof Rooth (2012). “Revealing taste-based discrimination in hiring: a correspondence testing experiment with geographic variation”. In: *Applied Economics Letters* 19.18, pp. 1861–1864.
- Dancygier, Rafaela M and David D Laitin (2014). “Immigration into Europe: Economic discrimination, violence, and public policy”. In: *Annual Review of Political Science* 17, pp. 43–64.
- Darolia, Rajeev et al. (2016). “Race and gender effects on employer interest in job applicants: new evidence from a resume field experiment”. In: *Applied Economics Letters* 23.12, pp. 853–856.
- Deming, David J et al. (2016). “The value of postsecondary credentials in the labor market: An experimental study”. In: *American Economic Review* 106.3, pp. 778–806.
- Dinesen, Peter Thisted and Kim Mannemar Sønderskov (2015). “Ethnic diversity and social trust: Evidence from the micro-context”. In: *American Sociological Review* 80.3, pp. 550–573.
- Gaddis, S Michael (2017). “How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies”. In: *Sociological Science* 4, pp. 469–489.
- Gaddis, S Michael and Raj Ghoshal (2015). “Arab American housing discrimination, ethnic competition, and the contact hypothesis”. In: *The ANNALS of the American Academy of Political and Social Science* 660.1, pp. 282–299.
- Gerber, Alan S and Donald P Green (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.
- Guryan, Jonathan and Kerwin Kofi Charles (2013). “Taste-based or statistical discrimination: the economics of discrimination returns to its roots”. In: *The Economic Journal* 123.572, F417–F432.
- Heckman, James J (1998). “Detecting discrimination”. In: *Journal of economic perspectives* 12.2, pp. 101–116.
- Jensen, Tina Gudrun et al. (2012). *Måling af diskrimination på baggrund af etnisk oprindelse*. SFI-Det nationale forskningscenter for velfærd.

- Kaas, Leo and Christian Manger (2012). "Ethnic discrimination in Germany's labour market: a field experiment". In: *German economic review* 13.1, pp. 1–20.
- Lancee, Bram et al. (2017). *Cultural distance and ethnic discrimination in hiring behaviour. Results from a cross-national field experiment*.
- Midtbøen, Arnfinn H (2016). "Discrimination of the second generation: Evidence from a field experiment in Norway". In: *Journal of International Migration and Integration* 17.1, pp. 253–272.
- Mudde, Cas (2013). "Three decades of populist radical right parties in Western Europe: So what?" In: *European Journal of Political Research* 52.1, pp. 1–19.
- Neumark, David (2018). "Experimental research on labor market discrimination". In: *Journal of Economic Literature* 56.3, pp. 799–866.
- Pager, Devah and Hana Shepherd (2008). "The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets". In: *Annu. Rev. Sociol* 34, pp. 181–209.
- Pager, Devah and Bruce Western (2012). "Identifying discrimination at work: The use of field experiments". In: *Journal of Social Issues* 68.2, pp. 221–237.
- Phelps, Edmund S (1972). "The statistical theory of racism and sexism". In: *The american economic review*, pp. 659–661.
- Riach, Peter A and Judith Rich (2002). "Field experiments of discrimination in the market place". In: *The economic journal* 112.483, F480–F518.
- (2004). "Deceptive field experiments of discrimination: are they ethical?" In: *Kyklos* 57.3, pp. 457–470.
- Rydgren, Jens (2008). "Immigration sceptics, xenophobes or racists? Radical right-wing voting in six West European countries". In: *European Journal of Political Research* 47.6, pp. 737–765.
- Sen, Maya and Omar Wasow (2016). "Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics". In: *Annual Review of Political Science* 19, pp. 499–522.
- Sidanius, Jim and Felicia Pratto (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

- Simonsen, Kristina Bakkær (2017). “Integration af indvandrere i Danmark.” In: *Politica: Tidsskrift for Politisk Videnskab* 49.3.
- Slot, Line Vikkelsø (2008). *Praktikpladser og formidlingspraksis: praktikkonsulentens i spændingsfeltet mellem virksomhedskrav og hensynet til ikke-diskrimination*. Rockwool Fondens Forskningsenhed.
- Zschirnt, Eva (2016). “Revisiting ethics in correspondence testing” . In: *NCCR-on the move, working paper series* 8.
- Zschirnt, Eva and Didier Ruedin (2016). “Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015” . In: *Journal of Ethnic and Migration Studies* 42.7, pp. 1115–1134.

Supplementary material: Experimental evidence of discrimination in the labour market

Appendix A

In this appendix, we describe our approach to selecting the names of the job candidates that were used in the experiment. The names were chosen from a dataset comprising the number of Danish citizens with a given first or last name. The popular ethnic majority first names were all among the top 20 most popular male and female names. The ethnic minority first names all figure in a dictionary of Middle Eastern first names that are in frequent use in Denmark (Meldgaard, 2005), and were identified from among the 20 most popular Middle Eastern names in use in Denmark. Furthermore, the last names were chosen from among the most popular Danish-sounding and Middle Eastern-sounding last names. Finally, we identified low-SES names based on a population index. We constructed the SES index using three items: average crime rate, unemployment rate, and annual income. The crime rate is a measure of the proportion of the population with a given name who have been incarcerated within the last five years. The unemployment statistic is a measure of the proportion of people with a given name who receive unemployment benefits or have been unemployed for at least six months. Income is the average annual income earned by those bearing a given name. All items were scored from 1 to 8, where a higher score indicates lower status (with income scored in reverse order). The selected low-SES names all belong to individuals in the bottom decile of the SES index with an average age between 20 and 40.

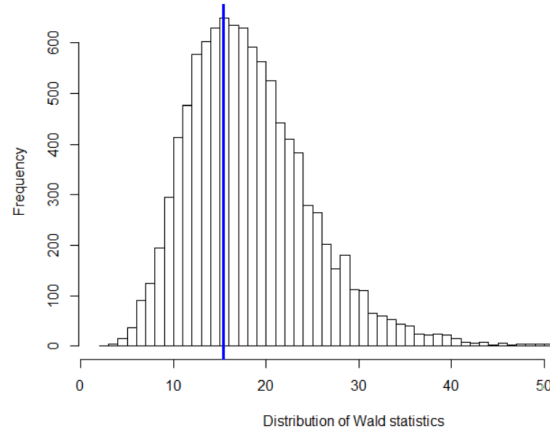
Table 1. Names of job applicants

Majority names					Minority names		
<i>Low SES name</i>		<i>Popular name</i>		-	-	-	-
<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>	<i>Last names</i>	<i>Male</i>	<i>Female</i>	<i>Last names</i>
Jimmi	Bonnie	Peter	Anna	Nielsen	Ali	Fatima	Ahmed
Sonny	Jennie	Michael	Mette	Jørgensen	Mohammad	Fatma	Sahin
Ricky	Jennifer	Martin	Anne	Pedersen	Ahmad	Aisha	Osman
Kenny	Belinda	Christian	Maria	Andersen	Mustafa	Hatice	Ismail
Ronnie	Michelle	Søren	Helle	Christensen	Ibrahim	Iman	Abdi
Nicky	Stella	Lars	Hanne	Rasmussen	Omar	Zainab	Farah
					Hassan	Amal	Yildiz
					Abdul	Amira	Ali

Appendix B

In this appendix, we test whether the chance of receiving a call-back is affected by the application process. We regress the outcome variable (call-backs) on three variables related to the application process: which application employers received first (the elapsed time between the two applications was 2–3 days), which of 16 unique applications was used, and which of two possible CVs was used. We calculate the heteroscedasticity robust Wald statistic for the hypothesis that all the coefficients on the variables are zero (blue line). In order to calculate the related p-value, we use randomization inference to create a reference distribution of Wald statistics from 10,000 permutations. The results show no reason to reject the null hypothesis ($p = .63$).

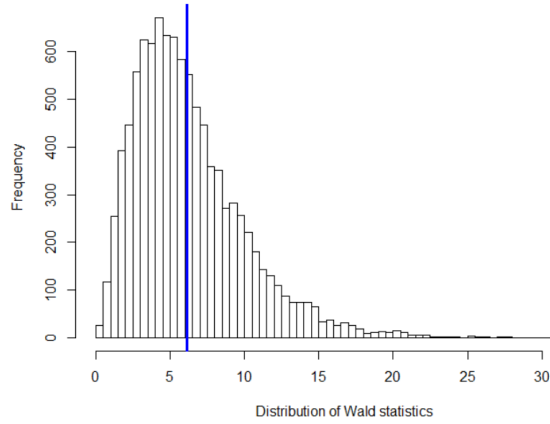
Figure B. Balance check



Appendix C

In this appendix, we perform a balance check to assess whether the random assignment of gender and SES appear to be reliable. We test whether observed covariate imbalances are larger than expected from chance alone. The covariates include five dummy variables: size (more or less than 50 employees), sector (public or private), education (if tertiary education was required), language requirements (if the job required good language abilities) and customer contact (frequent customer contact). We calculate the heteroscedasticity-robust Wald statistic for the hypothesis that all the coefficients on the variables are zero (blue line). In order to calculate the related p-values, we use randomization inference to create a distribution of 10,000 Wald statistics. As Figure C1 and Figure C2 show, the results does not indicate a reason to reject the null hypothesis ($P = .43$ and $p = .92$).

Figure C. Balance check for assignment of gender



Appendix D

In this appendix we show the descriptive statistics of the sample across the six occupational categories and the various occupations (23 occupations in total). The table includes the number of applications that were sent for each occupation (N), the call-back rates for both majority and minority candidates, and the gender composition in the specific occupation. Gender composition is based on data from the Danish National Centre for Social Research (Larsen et al., 2016) and denoted *Balanced* if the share of males in a given occupation is between 20 and 80 percent, *Male* if the share of females is < 20 percent, and *Female* if the share of males is < 20 percent.

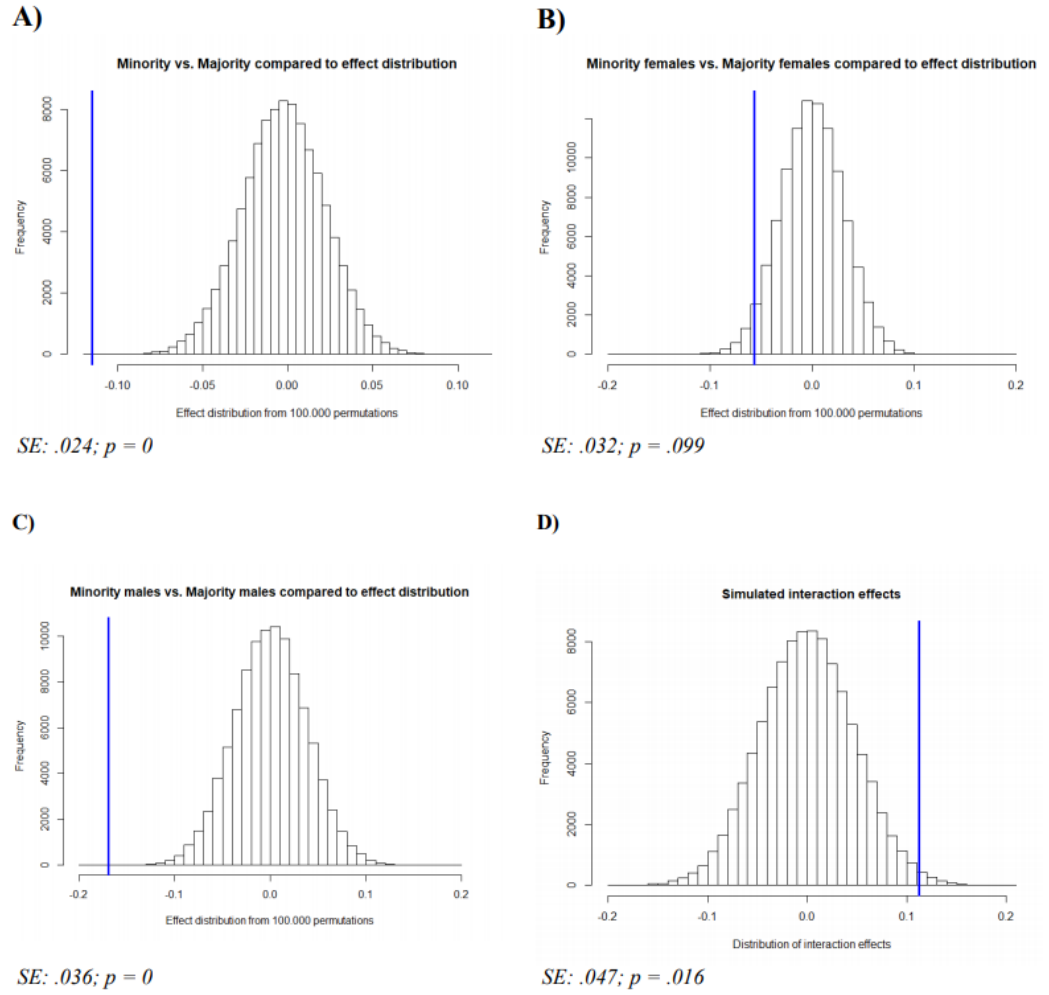
Table D. Descriptive statistics by occupational categories

<i>Occupational categories and occupations</i>	<i>N</i>	<i>Call-back majority</i>	<i>Call-back minority</i>	<i>Ratio</i>	<i>Private sector jobs</i>	<i>Gender composition</i>
Administrative support	66	.136	.106	1.29	56	
Accountancy	8	.125	.000		6	Balanced
Administrative assistant	20	.150	.150		16	Balanced
IT-support	15	.133	.133		13	Male
Receptionist & secretary	23	.130	.087		21	Female
Construction & Extraction	81	.358	.148	2.42	79	
Building and construction	4	.750	.000		4	Male
Carpenter	9	.333	.222		9	Male
Gardener	1	.000	.000		1	Balanced
Industrial production	7	.429	.143		6	Balanced
Mechanic	6	.500	.333		6	Male
Driver	25	.520	.200		25	Male
Warehouse	29	.138	.069		28	Male
Education	55	.382	.255	1.50	6	
Pedagogue	29	.310	.207		3	Female
Teacher	26	.462	.308		3	Balanced
Health care	60	.433	.400	1.08	7	
Nurse	14	.385	.462		3	Female
Rehabilitation assistant	17	.294	.353		1	Balanced
Social and health care assistant	22	.455	.364		1	Female
Social worker	8	.750	.500		1	Female
Retail	98	.306	.173	1.76	91	
Cleaning	17	.294	.059		14	Female
Restaurant & Hotel	19	.263	.158		19	Balanced
Retail and service	62	.323	.210		58	Balanced
Sales & related	40	.475	.350	1.36	39	
Communication and marketing	6	.333	.333		6	Balanced
Salesman	12	.250	.083		11	Balanced
Telemarketing	22	.636	.500		22	Balanced
Total	400	.335	.220	1.52	278	

Appendix E

In this appendix, we visualise the approximate permutation tests for the effect of the ethnic trait among all candidates, female candidates, and male candidates, respectively. Furthermore, we visualise the approximate permutation test for the interaction between gender and ethnicity. We use the R package *randomizr* to perform these tests. The fact that each unit received both treatment and control increases precision. Rather than comparing call-backs across a heterogeneous collection of employers, we increase statistical power by controlling for employer-level heterogeneity. We assess statistical uncertainty around each estimate by running 100,000 permutations with block random assignment of ethnicity at the job level under the sharp null assumption of no effects for all units. This corresponds to an OLS regression with fixed effects. While the ethnic indicator variable is manipulated within jobs, gender is manipulated across jobs. Hence, the interaction model between ethnicity and gender is analysed assuming constant, additive treatment effects with block random assignment of ethnicity and cluster random assignment of gender at the job level. Hence, Figure E1 presents randomization inference for the ethnic difference among A) all applicants, B) female applicants and C) male applicants. Finally, D) presents the interaction between minority status and gender

Figure E. Hypothesis tests



Appendix F

In this appendix we show three specifications of the interaction model. Column (1) reports the covariate-unadjusted model that is visualized in Figure 2 in the paper. Column (2) reports the covariate-adjusted model. Finally, to compare results with an OLS regression model, column (3) reports the covariate-unadjusted model with job fixed effects and cluster corrected standard errors.

The rationale behind including covariates is to reduce disturbance variability, but the inclusion of covariates does little to improve the precision of our estimated treatment effects (in fact, the standard errors increases slightly in the adjusted model). The covariates include five dummy variables: size (more or less than 50 employees), sector (public or private), education (if tertiary education was required), language requirements (if the job required good language abilities) and customer contact (frequent customer contact). The fact that ethnicity is randomly assigned within occupations while gender is assigned between occupations implies that the effect of ethnicity is measured with more precision than the effect of gender. We account for this in model (1) and model (2) by using randomization inference with the same randomization scheme – gender is clustered at the job level and ethnicity block randomized at the job level – with 100,000 iterations and the effect imputed for each iteration to obtain standard errors.

Figure F (i). Interaction between minority status and gender

Variables	(1)	(2)	(3)
Minority	-.169 (.033)	-.169 (.034)	-.169 (.034)
Female	-.056 (.044)	-.068 (.047)	-.056 (.047)
Minority * Female	.112 (.047)	.113 (.047)	.112 (.046)
Covariate adjustment	No	Yes	No
N = 400			

Interaction between minority status and SES of the majority applicant

Column (1) reports the covariate-unadjusted model, while column (2) reports the covariate-adjusted model (using the same covariates as in Table F1 above). We estimate uncertainty in model (1) and model (2) by using randomization inference with the same randomization scheme – SES is clustered at the job level and ethnicity block randomized at the job level – with 100,000 iterations and the effect imputed for each iteration to obtain standard errors. Finally, to compare results with a conventional OLS regression model, column (3) reports results from a regression of a call-back dummy on the Minority-indicator, the SES-indicator, and the interaction between the two. Standard errors

are cluster corrected.

Figure F (ii). Interaction between minority status and SES of the majority applicant

Variables	(1)	(2)	(3)
Minority	-.124 (0.033)	-.121 (.043)	-.124 (.033)
Popular name	.043 (0.047)	.05 (.044)	.043 (.047)
Minority * Popular name	-.018 (.048)	-.012 (.062)	-.018 (.047)
Covariate adjustment	No	Yes	No
N = 400			

Appendix G

To enhance transparency, the hypotheses were preregistered before any data was collected. The preregistration can be found at <http://egap.org/> with ID 20150930AA. The preregistration document contains a very brief introduction to the study, its design, the three hypotheses and the outcomes of interest. It should be noted that the preregistration document omits important information such as the sampling frame, attrition criteria, a specific plan with technical details for the data analysis, and explicit power calculations. This is all information that it would of course have been appropriate to include.

Appendix H

This appendix reports call-backs across gender, ethnicity and SES of the majority candidates and the difference in differences.

Figure H: Call-backs across ethnicity, SES and gender

	Males		Females	
	<i>Call-backs when majority candidate has a popular name</i>	<i>Call-backs when majority candidate has a low-SES name</i>	<i>Call-backs when majority candidate has a popular name</i>	<i>Call-backs when majority candidate has a low-SES name</i>
Majority	.409	.309	.292	.317
Minority	.236	.144	.269	.231
Difference	.173	.165	.022	.086
N	110	97	89	104
Difference -in-differences	0.008 (0.07)		-0.064 (0.063)	

Alike but different: How cultural distinctiveness shapes immigrant-origin minorities' access to the labour market

Malte Dahl*

Under review in Ethnic and Racial Studies

Abstract

Does cultural dissimilarity explain discrimination against immigrant-origin minorities in the labour market? I conducted a factorial field experiment ($N = 1,350$) to explore how explicit group cues trigger differential treatment and whether individuating information that counters cultural-based stereotypical representations mitigate discrimination. Employers were randomly assigned a job application with a putative female ethnic majority or immigrant-origin minority alias and CV photographs portraying the minority candidate with or without a headscarf – perhaps the quintessential marker of Muslim identity. Moreover, half the job applications conveyed information intended to reduce cultural distance by indicating a liberal lifestyle and civic participation. The results demonstrate that immigrant-origin women are significantly less likely to receive an invitation to a job interview, especially if they also wear a headscarf. Contrary to expectations, the differential treatment is not moderated by the individuating information in the applications. This indicates that the differential treatment is persistent and also targets immigrant-origin minorities who have acquired soft skills and signals cultural proximity.

*This project has received funding from the Danish Institute for Human Rights. I am grateful to Michael S. Gaddis, Arnfinn Midtboen, Peter Thisted Dinesen, Jonas H. Hansen, the participants at the IMISCOE Annual Meeting 2018 and the participants in the IWG seminar at UC Berkeley for helpful comments on this paper. Kasper Arabi and Lotte Andersen deserve special thanks for their invaluable research assistance.

Introduction

Addressing the religious and cultural heterogeneity stemming from the increasing numbers of immigrant-origin minorities is a key challenge today and in the foreseeable future to most Western societies. Understanding and countering discrimination against immigrant-origin minorities is an important part of this process. Consistent evidence of widespread discrimination indicates that members of immigrant-origin minority groups face significant barriers in their access to labour markets (Zschirnt and Ruedin 2016; Baert 2018; Quillian et al. 2017). Studies also find evidence indicating that employers discriminate against second-generation immigrants, thus suggesting that ascribed group categories conveyed by names or psychical appearance represent barriers to immigrant-origin minorities who have acquired linguistic fluency and formal education (Midtbøen 2014; Carlsson 2010). This is not only problematic from an economic perspective but also poses an impediment to integration and social cohesion more generally by undermining the belief in meritocratic principles and equal opportunity (Oskooii 2018; Adida, Laitin, and Valfort 2016).

A burgeoning body of literature examines the factors underlying negative attitudes to and discrimination against immigrant-origin groups. One prominent stream of scholarship contends that such out-group biases are related to perceptions of cultural distinctiveness of minority groups (Sniderman, Hagendoorn, and Prior 2004; Hainmueller and Hangartner 2013). Specifically, it is suggested that attitudes and behaviours towards out-groups are shaped by cultural, ethnic and religious differences (Sides and Citrin 2007; Konitzer et al. 2019), and that majorities rank out-groups according to perceptions of cultural distance from the majority group (Hainmueller and Hopkins 2014; Ostfeld 2017). Moreover, survey experiments find that signals of individual immigrants' level of assimilation into cultural norms and traditions are an important indicator of majorities' attitudes towards them (Hopkins 2015; Iyengar et al. 2013). While this evidence indicates that perceptions of cultural distance are crucial to explaining negative sentiments towards immigrant-origin minorities, less attention has been paid to how such perceptions translate into actual discriminatory behaviour.

A related outstanding question concerns the way individuating information that counters stereotypical representations can serve to mitigate discriminatory behaviour. The bulk of experimental

research attempting to elicit how information affects discrimination in the hiring process has manipulated the quality of credentials in the job applications (i.e. hard skills such as education or years of labour market experience) (Bertrand and Mullainathan 2004; Oreopoulos 2011; Nunley et al. 2016). However, little attention has been paid to the question of whether discrimination is related to perceptions of cultural distinctiveness. Maybe because it is typically classified as non-productive, job applicants' displays of cultural signals and lifestyle markers (soft skills) have received little empirical attention (Rivera 2012). The question of how soft skills affect discrimination against immigrant-origin applicants has a bearing on debates concerning the importance of assimilation for labour market integration (Koopmans 2016) and to what extent immigrant-origin minorities can counteract employer discrimination by accumulating skills.

This paper addresses these questions by theorizing that (i) discrimination against individual out-group members increases with the immediate salience and cultural distance of the out-group, while (ii) information that contradicts cultural-based stereotypes by signaling adherence to cultural norms can counter stereotypes and mitigate discrimination. This is tested in two ways using a pre-registered correspondence experiment in the Danish labour market.¹ First, to measure how signals of cultural distance shape discrimination, the experiment's main treatment is *explicit* group categories indicated by names and CV photographs. Applicants are endowed with either a traditional Danish (majority) alias or a non-Western (minority) alias. Moreover, the CV photographs enable me to examine an essential component in perceptions of cultural distance, namely Muslim affiliation. While this has been a topic of interest in previous field experimental research on discrimination, identifying a religious effect is not straightforward due to concerns about authenticity and contextual factors.² I overcome this methodological challenge by taking advantage of the fact that it is common in the Danish labour market to include a personal photograph in a CV. Hence, it is feasible to

¹ The study was pre-registered at EGAP.org (ID: 20170412AA).

² A few studies manipulate explicit affiliation to religious organisations in résumés (Pierné 2013; Wright et al. 2013). While this is a feasible strategy in some contexts, this type of information is not an integral part of job applications in most labour markets, and therefore threatens authenticity and real-world relevance. Other studies use names as signals for different religious affiliations among the same ethnic group (Adida, Laitin, and Valfort 2010), but this is not an option in countries where religion and ethnic affiliation is often intertwined.

randomly manipulate CV photographs and treat employers with applicants wearing headscarves.³ Surprisingly, given the pertinence of this cultural and religious symbol, the effect of wearing a headscarf has not been the subject of extensive study in European correspondence experiments on labour market discrimination, with the exception of Weichselbaumer (2016). Secondly, I examine the importance of *implicit* signals of cultural distance by providing individuating information in the job applications intended to tap into perceptions of cultural proximity. Specifically, I randomly assign information related to cultural dimensions – comprised here by applicants’ implicit support for a liberal lifestyle and civic engagement.

In total, 1,350 generic applications were sent to job openings listed on the largest online job portal in Denmark. Each job opening was randomly assigned a job application from either a candidate with (i) a traditional Danish name, (ii) a Middle Eastern name or (iii) a Middle Eastern name and who wore a headscarf. A picture of the same person was used in every CV. Moreover, half of the submissions included additional information in the applications and CVs (resulting in a 3x2 factorial design). The jobs spanned both public and private sector jobs in order to assess sector differences. Applicants were highly qualified for the jobs applied for and always possessed relevant experience and the required educational background.

The paper reports three main findings. First, in line with previous studies, I find overall evidence of discrimination against immigrant-origin minorities in the labour market. In line with the expectations, the results demonstrate a clear hierarchy when comparing the number of invitations received to job interviews. Majority applicants received 4.6 percentage points more invitations to attend interviews than minority applicants without a headscarf. When the minority applicants wore a headscarf, this difference increased to 11.5 percentage points, implying a significant penalty of wearing the headscarf of around 7 percentage points. Second, contrary to the expectations, the additional implicit information that contradicted cultural stereotypes did not moderate differences between the ethnic majority and the two immigrant-origin minority groups. This indicates that group-based discrimination is persistent and not easily altered by individuating information. In

³ Initial interviews with women wearing headscarves revealed that women often include pictures in their applications so as not to surprise employers at job interviews. They also argued that social media and LinkedIn profiles would ‘reveal’ the fact that they wore headscarves anyway.

other words, individuals with an immigrant-origin minority background remains to be discriminated against, also when they signal cultural similarity. Third, these differences are both evident in public and private sector jobs and consistent across occupational categories, thus contradicting the common notion that the diversity policies and standardised application templates in use in public sector jobs eliminate discrimination.

Hiring discrimination against immigrant-origin minorities

Findings from studies on ethnic discrimination in the labour market are remarkably consistent, with overwhelming support for the proposition that immigrant-origin minority applicants are discriminated against (Birkelund, Heggebø, and Rogstad 2017; Adida, Laitin, and Valfort 2010; Bursell 2014; Arai, Bursell, and Nekby 2016; Blommaert, Coenders, and Van Tubergen 2013). For an overview and meta-analysis, see Baert (2018) and Zschirnt and Ruedin (2016). The literature on labour market discrimination indicates that the prevalence of ethnic discrimination is dependent on the minority groups under examination (so-called ethnic hierarchies). Everything else being equal, members of particular ethnic minority groups are the targets of more severe discrimination than others (Lancee 2019; Booth, Leigh, and Varganova 2012; Fibbi, Lerch, and Wanner 2006).⁴ This finding is also replicated in other empirical contexts.⁵ Moreover, there is ample evidence that negative sentiment towards immigrants is conditional on their country of origin (Brader, Valentino, and Suhay 2008; Hainmueller and Hopkins 2015).

What explains the apparent singling out of particular minority groups? One prominent theory departs from the premise that majorities evaluate ethnic or racial out-groups based on their perceived adherence to norms. The greater the perceived dissimilarity of immigrant-origin groups to the majority's general cultural grounds, the more likely they are to be met with hostility (Schildkraut 2010; Ostfeld 2017; Hainmueller and Hopkins 2014). Members of racial or ethnic out-groups are thus evaluated differently according to the ways in which these groups are perceived to pose a

⁴ Although there are exceptions, e.g. Andriessen et al. (2012) and McGinnity and Lunn (2011)

⁵ For example, ethnic hierarchies are found in research on how likely local election officials are to respond to voters' inquiries on where to vote (Hughes et al. 2017) or in decisions of naturalization (Hainmueller and Hangartner 2013).

cultural-symbolic threat to the nation's religion, ethnicity, identity and culture (Brader, Valentino, and Suhay 2008; Harell et al. 2012; Adida, Laitin, and Valfort 2016).⁶ That perceptions of cultural distinctiveness shape minority-groups' access to the labour market is compatible with empirical evidence from recent studies on labour market discrimination. For example, a cross-national field experiment by Lancee (2019) indicates that discrimination correlates with the extent to which minority groups' countries of origin deviate from the receiving society with regard to secular and emancipative values or the size of their Muslim populations. Furthermore, there is evidence indicating that religious affiliation, specifically being Muslim, amplifies discrimination (Ahmed 2010; Adida, Laitin, and Valfort 2016; Weichselbaumer 2016; Pierné 2013; Wright et al. 2013). For example, in a series of studies in a French context, Adida, Laitin, and Valfort (2016) shows that immigrant-origin minorities experience significantly more discrimination when they are Muslim. In contrast to this experimental evidence, some scholars have argued that once sociocultural variables such as language proficiency, interethnic social ties and gender values are accounted for, differences in labour market participation is limited, and the link between experienced discrimination and unemployment disappears (Koopmans 2016).

The link between cultural distance and discrimination can be understood within the framework of traditional economic models of discrimination that typically outline two rivalling theories. In the *taste-based* discrimination models, discrimination is caused by irrational distastes towards certain groups, resulting in preferential hiring and wage differentials (Becker 2010). Although there is no accounting for how people form unfavourable attitudes in the original formulation of the theory, it is plausible that dis-tastes towards out-groups correlate with perceptions of cultural distance. Theories of *statistical* discrimination contend that discriminatory behaviour is a result of an employer optimisation strategy in which employers use group cues to make decisions under conditions of uncertainty (Aigner and Cain 1977; Phelps 1972). The degree to which candidates appear socially or culturally distant may be relevant for productivity or correlated with unobservable variables that matters for productivity. Hence, the notion that cultural distance matters is seemingly consistent

⁶ Another example is Sniderman, Hagendoorn, and Prior (2004), who provide experimental evidence indicating that opposition to immigration is rooted in cultural concerns and conclude that 'fitting in culturally promotes significantly more support for it [immigration] than fitting in economically'.

with both notions.

Explicit and implicit signals of cultural distance and why they matter

To explore the effect of an *explicit* signal of cultural distinctiveness, the present study examines the effect of having a Middle Eastern name alone and in combination with wearing a Muslim headscarf. It is important to note that, while wearing a headscarf can be seen as an examination of Muslim affiliation, it is evidently not exclusively a signal of religion (Butler and Tavits 2017). Using photographs of women wearing headscarves is a compound treatment that is likely to induce perceptions of cultural distance more generally. As such, the headscarf serves as a signal that is attributed numerous symbolic meanings.

There are reasons to expect that wearing a headscarf increases perceptions of cultural distance and negatively affect employers' evaluations of candidates. Sniderman, Hagendoorn, and Prior (2004) highlights three factors that conditionally strengthen the feeling of cultural distance: (i) how easily immigrants can be identified from native citizens in everyday life; (ii), the salience of immigration topics in public; and (iii), the extent to which the migrant group is perceived as a group that is clearly distinguishable from the native and other migrant groups. The headscarf arguably satisfies all three conditions.

The headscarf is often considered a signal of intolerance and as indicating an unwillingness to integrate (Byng 2010; Shadid and Van Koningsveld 2005). Studies show that natives often find that the headscarf conflicts with egalitarian gender roles and liberal social values (Bowen 2007; Helbling 2014). Helbling (2014) has demonstrated how natives hold significantly more negative views towards the headscarf than towards Muslims in general in six European countries. Moreover, immigrant-origin minorities from Muslim countries tend to be perceived as culturally distant from Western European countries in terms of cultural values (Norris and Inglehart 2012; Statham and Tillie 2016). Finally, in a field experiment in the German labour market, Weichselbaumer (2016) finds that minority women wearing a headscarf were much less likely to receive an invitation to a job interview.

To explore the importance of *implicit* signals of cultural distance, employers were randomly assigned to either a baseline job application or an application that included individuating information that taps into dimensions of cultural proximity. There are good reasons to expect that information that counters cultural-based stereotypes matter. In general, similarities in experience and culture have been found to be crucial in studies on evaluations of job candidates (Rivera 2012) and students (Erickson and Shultz 1982). In fact, Rivera (2012) suggests that employers' cultural concerns outweigh concerns about candidates absolute productivity. It is also worth mentioning that a cultural match have been found to be highly salient to employers when hiring immigrant-origin minorities in Denmark (Slot 2008).

Previous experiments have studied whether higher-quality applications reduce the differential treatment faced by racial or ethnic minorities. While these studies generally find no effect of improving applicants' credentials such as education or years of labour market experience (Nunley et al. 2016; Bertrand and Mullainathan 2004; Oreopoulos 2011), there are two notable exceptions. In a correspondence study in Germany, Kaas and Manger (2012) found that discrimination against Turkish job applicants was eliminated when a reference letter, containing indirect information about conscientiousness and agreeableness, was randomly included. Agerström et al. (2012) found that when candidates were presented as both more competent and warm, the relative difference in callbacks between minorities and majority candidates was reduced compared to a situation where candidates appeared only as competent.⁷ These findings suggest that information on soft skills may affect employers' evaluations of immigrant-origin candidates. In other words, while there seems to be limited pay-off from increasing 'hard skills', individuating information about social and cultural proximity may detach immigrant-origin minorities from cultural stereotypes and positively affect their hiring prospects.

For the purpose of the present study, it is crucial to consider values that can be expected to be salient to employers when making hiring decisions as well as information that is realistic to include

⁷ Warmth is signalled by described himself as a 'warm and social person who gets along great with others' both at work and elsewhere'

in job applications. In order to tap into perceptions of cultural stereotypes, I rely on the work of Norris and Inglehart (2012) on how migrants carry their culture with them, and to what extent they acquire the culture of their new setting. Specifically, the information-treatment taps into the two dimensions of *Secular-rational* values and *Self-expression values*; two domains where non-Western countries deviate significantly from Scandinavian countries.⁸ Information related to these dimensions is signaled by indicating active civic participation and liberal values that contradicts a traditional lifestyle (see operationalization in the Design section). While this type of information should have no or only small positive effects on majority applicants' chances of receiving a callback, employers might consider it to be an important source of uncertainty in relation to hiring members of an immigrant-origin minority group.

Context

Since the late 1960s, Denmark has experienced a rapid increase in the number of non-Western immigrants and descendants, with the largest groups being from Turkey, Lebanon, Pakistan and Iraq. Non-Western immigrants and their descendants now comprise approximately 8.5% of the total Danish population (DST 2018). The integration of non-Western immigrants and their descendants is a contested and important political issue in Denmark (Simonsen 2017). Discussions about the economic and cultural integration of immigrants are often motivated by the fact that minorities of Middle Eastern origin fare worse on a number of socioeconomic indicators (DST 2018).

The cultural aspect of integration is a key factor in these discussions. For example, a representative poll from 2011 indicated that 92% of Danes thought that immigrants and their descendants should live in accordance with Danish norms. Evidence from quantitative and qualitative work supports the notion that cultural aspects are important, indicating that employers perceive language and cultural issues as barriers when hiring ethnic minorities (Slot 2008). The salience of the topic is also reflected in discussions about religious symbols, and it has been recently suggested, albeit

⁸ Self-expression values include social toleration, public expression and an aspiration to liberty. Secular-rational values place less emphasis on traditional family values and religion.

not widely supported, banning headscarves in public institutions. A 2017 poll revealed that half the population indicated that it would bother them to be serviced by a cashier wearing a headscarf at work (Ramboell, 2017).

A distinctive feature of the Danish labour market is the large public sector, which is one of the largest in Europe (Economic Cooperation and Development 2017). Some evidence indicates that there is reason to believe that the public sector is less plagued by discrimination than the private sector (Reskin 2011). Job advertisements in the public sector explicitly invite minorities to apply, and candidates applying for positions in the public sector are required to use job templates, which have been seen as a means to counter discrimination (Midtbøen 2016). However, recent research questions this notion (Villadsen and Wulff 2018; Dahl and Krog 2018), which emphasizes the importance of examining discrimination in both the public and private sectors.

Experimental design

Main treatments: Names and photographs

Paralleling previous field experiments, the present study relies on names and photographs as group signifiers. All the names were among the most commonly used names in Denmark (Statistics Denmark 2015). Furthermore, the study included two pools of surnames with the most commonly used traditional Danish-sounding and Middle Eastern-sounding surnames that were randomly paired with the pool of first names. In total, the study relied on 160 different combinations of first and last names. See a list of names in Appendix A. Since the most common Middle Eastern-sounding names are quite distinct, it can be assumed that employers are able to differentiate between the ethnic traits.

The CVs included a photograph of the applicant in a black and white format in the CVs (see details in Appendix A). To ensure that the only difference was the headscarf and that the applicants were equally attractive, I relied on pictures of the same person. The use of CV photographs of women wearing headscarves raises important questions about the degree of ‘fieldness’ related to the authenticity of the treatment: does the use of photographs resemble the reality of the labour

market? While it is quite common to attach photographs to job applications in the Danish labour market – especially in certain types of jobs – these are rarely explicitly requested.⁹ It could be argued that women wearing headscarves could avoid including pictures in their applications and thereby circumvent negative effects. However, interviews with women wearing headscarves revealed that it is common practice to include a photograph in order to avoid “surprising” the employer at a possible interview. They also argued that employers could retrieve the same information on social media or LinkedIn, leaving no reason to avoid including a photograph. In other words, while the use of CV photographs is not necessary, it is common practice. It should also be noted that I did not receive comments on the pictures for any of the candidates in rejections or invitations to interviews.

Secondary treatment: implicit stereotype-reducing information

Half of the job advertisements were randomly assigned job applications and CVs with additional individuating information related to cultural proximity. In Danish job applications, it is common to add a brief section outlining the applicant’s leisure activities, which I utilise in this experiment. The information was randomly included in half of the job applications and CVs by adding the information set out in Table 1. To indicate civic participation and majority contact, the CV included the information that the applicant participated as a board member in her housing organization¹⁰ and participated in sports. Immigrant-origin minorities, especially women, are significantly less likely to be members of sports clubs compared to the average Danish citizen (Agergaard 2008; Ministry of Employment 2018) – this is especially true with regard to handball, a highly popular sport in Denmark (Ejsing 2014). Thus, to counter a potential stereotype, the application indicated that the applicant had a connection to the local community by participating in her local sports

⁹ One way to identify the effect of religious affiliation is to manipulate explicit affiliation to religious organisations in CVs (Pierné 2013; Wright et al. 2013). Since this type of information is not an integral part of job applications in most labour it threatens authenticity and real-world relevance. Other studies have used names as signals for different religious affiliations among the same ethnic group (Adida, Laitin, and Valfort 2010), but this is not an option in the Danish context where religion and ethnic affiliation are intertwined.

¹⁰ Cooperative social housing organisations are self-governing associations, with a management council controlled entirely by residents through a policy of tenant democracy. Cooperative housing is very common in the Danish housing market, which has 540,000 cooperative housing units.

club. Furthermore, to signal lifestyle markers that indicated a liberal lifestyle, it was stated that the applicant loved travelling and had a partner with a majority name.

Table 1. Individuating information randomly assigned to job application and CV

‘In my spare time I engage in my local community. For example, I am active in my local handball club and attend practice twice a week’

‘I like to travel and see the world, and especially I love to travel with my boyfriend Martin’

‘– I served as a board member for three years in my then co-operative society’ [Added to CV]

Constructing applications and CVs

All applications comprised a letter of application and a CV. The applications were constructed using a framework of four paragraphs: introduction, motivation, experience and personal interests. The applicants had high-quality CVs, always possessing the necessary educational background and relevant experience from previous jobs. The fictitious applicants had conducted all of their schooling in the Danish educational system and had experience from relevant jobs in the Danish labour market. Additionally, the applications were written in flawless Danish.

The application templates were created prior to commencing data collection, and each template matched a specific type of job. If the job advertisement explicitly mentioned specific requirements, these were incorporated into the CVs (e.g., software skills, a driver’s licence). By finding a job and collecting descriptive data without knowing the applicant type (treatment or control), the research assistants avoided the risk of experimenter effects (unintentionally biasing the quality of the applications).

Outcome measure

The main outcome of interest is employer contact, measured as callbacks rather than actual job offers. A callback is here defined as a personalised phone or e-mail contact by a potential em-

ployer with an invitation to a job interview or a request for additional information. All replies were responded to within 24 hours by politely turning down the invitation for an interview.¹¹ Some research within other areas (such as communication between street level bureaucrats or politicians and citizens) indicates that the content of communication is an important dimension of discrimination (Einstein and Glick 2017; Hemker and Rink 2017). However, most callbacks are completely standardised, minimising the value added by including measures of content.

Randomisation details

By exposing employers to different manipulations of randomly assigned characteristics, the study disentangles how different aspects both alone and in combination affect employers' behaviour. The 3x2 factorial design resulted in six different conditions: for each of the three applicant types, additional information was either included or not. A block-randomised design was applied in order to randomly divide the sample into treatment and control groups across six occupational categories. The blocking incorporated covariates into the design and ensured that treatments were balanced.¹² By following the same randomisation procedure in the analysis, p -values and confidence intervals are based on randomisation inference (Gerber and Green 2012, p. 150).

Because the study includes three groups and use CV photographs, the matched-pairs design that is often applied in correspondence studies (i.e., sending two or more applications to the same job) is not well suited.¹³ Instead, each job opening received only one application. While this design in most situations implies a loss of statistical precision compared to the matched-pairs design, it also avoids some of the drawbacks related to matched-pairs design (Vuolo, Uggen, and Lageson 2018). Since each experimental unit in a matched-pairs design receives both treatment and control, the risk of invoking suspicion among employers would be greatly increased by the use of three nearly

¹¹ Ethical considerations are outlined in Appendix B.

¹² An advantage of block-randomised design is that it effectively reduces random differences between the treatment and control groups, and if covariates are prognostic of the outcome, it enables higher-powered comparisons by reducing baseline differences (Horiuchi, Imai, and Taniguchi 2007).

¹³ Matched-pairs designs typically have greater statistical power. However, the difference in sample size requirements between matched and unmatched designs becomes smaller as the overall level of callbacks increases and differences between the treatment and control groups increase. Hence, in this particular experiment, it has small consequences for statistical power.

identical applications – all of them including pictures – two of which were from ethnic minority candidates (Lahey and Beasley 2018).

The sample

The empirical analysis builds upon experimental data collected by two research assistants between February 2017 and September 2017. In total, 1,350 applications were sent in response to the same number of job openings. My research team and I sampled the jobs from the largest online job portal in Denmark, Jobindex.dk. To provide a comprehensive picture of potential discrimination, the research team applied for jobs across all Danish cities within six occupational categories. We excluded academic jobs and jobs that required highly technical expertise, since these would have required detailed knowledge of essential skills, specific applications, recommendations and lengthy CVs that was not possible to include in the generic applications used in this experiment.

Empirical results

Table 2 presents initial descriptive statistics broken down by the treatment conditions across occupations and sectors. Overall, the applicants received a callback in one out of every four jobs, which indicates that the applications were generally of high quality and that the labour market is characterised by a high demand for labour.¹⁴ Three results are especially worth noting. First, and most importantly, the share of callbacks to the three types of applicants differed substantially. The results provide evidence of bias, with the majority candidate receiving the largest share of callbacks (30.6%), with the minority candidate receiving the second-largest share (26%) and the minority candidate wearing a headscarf receiving a substantially lower share of callbacks (19.1%). The effect estimates (difference in means) and confidence intervals are depicted in Figure 1. Second, no overall positive effect was observed as a result of adding additional information to the applications. Third, there was considerable variation in the callback rates across occupational categories. Overall, the

¹⁴ Labour market competitiveness has been found to curb discrimination (Baert et al. 2015), although there is also evidence indicating otherwise (Carlsson, Fumarco, and Rooth 2018).

applicants received substantially higher shares of callbacks from jobs in health care, while the low number of callbacks for jobs within administrative support indicates a higher level of competition for these types of jobs. Additional descriptive results for workplace characteristics are reported in next section.

Table 2. Descriptive statistics

	Callbacks	<i>N</i>
Total	0.253	1350
By characteristic		
Majority	0.306	450
Minority	0.26	450
Minority + headscarf	0.191	450
By information		
Baseline application	0.259	675
Application w. individuating information	0.246	675
By occupation		
Administrative support	0.075	225
Construction & extraction	0.177	225
Retail	0.231	225
Sales & related	0.173	225
Education	0.284	225
Health care	0.573	225
By sector		
Public	0.384	429
Private	0.191	921

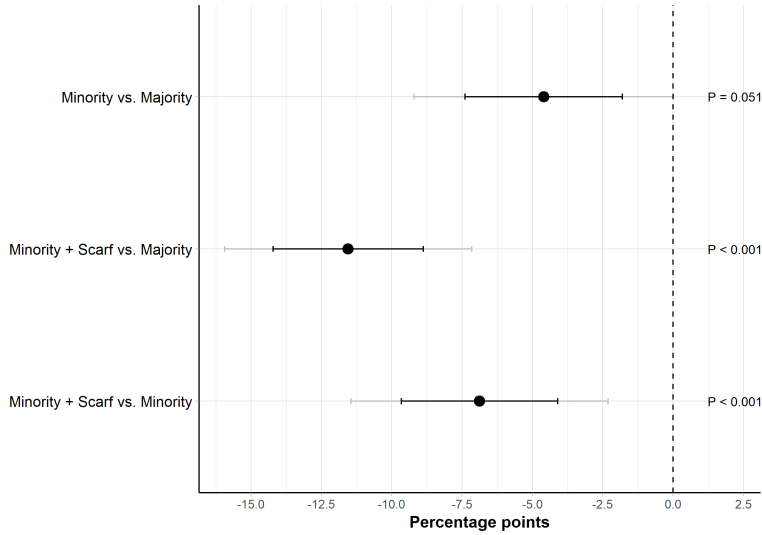
Note: The table shows callbacks and the number of jobs applied for by conditions.

Figure 1 presents the difference in means estimates and related measures of uncertainty when comparing callback rates across the group categories as well as p -values.¹⁵ The difference between majority and minority candidates without a headscarf is 4.6 percentage points and borderline significant ($p = 0.05$). The result implies that minority candidates would have to send out 18% more applications than majority applicants to receive the same number of invitations to job interviews.

¹⁵ Confidence intervals are constructed from an inverted hypothesis test by computing a full schedule of potential outcomes under the sharp null hypothesis of no effect for all units and re-drawing 100,000 experiments in order to calculate the p -values and confidence intervals related to the differences (Gerber and Green 2012).

This result is almost identical to those of a previous correspondence experiment in the Danish labour market that also explored callbacks across minority and majority female job applicants (Dahl and Krog 2018). While one should generally be cautious when comparing results from correspondence experiments directly – results are sensitive to context, design and the quality of the applications used – the experimental procedures across the two studies are highly comparable.¹⁶ This comparison shows almost identical effects of having a female minority name (4.6 percentage points and 5.6 percentage points respectively), supporting the validity of the findings. The close replication strengthens the belief in the precision of the findings. A difference of 18% is lower than obtained in comparable European correspondence experiments – especially those studying discrimination against Middle Eastern minorities. However, the few studies that explore discrimination against female job applicants (as compared to male applicants) generally indicate smaller levels of ethnic discrimination (Arai, Bursell, and Nekby 2016; Midtbøen 2016; Dahl and Krog 2018). Figure 1 also provides evidence on the importance of wearing a headscarf. The difference in callback rates increases to 11.5 percentage points when comparing callbacks between minority candidates wearing a headscarf and majority candidates. This represents a difference of almost 60% compared to the majority applicants’ rate of invitations to job interviews. Finally, the isolated effect of the headscarf can be estimated by comparing the callback rates between minority candidates with and without the headscarf of 7 percentage points, corresponding to a significant difference of 36%. In other words, similar to findings from the German labour market, wearing a headscarf substantially increases discrimination against the immigrant-origin minority group (Weichselbaumer 2016).

¹⁶ Both experiments build on roughly the same occupational categories and use the same outcome measure as well as almost identical applications and CVs, justifying their integration into a combined analysis.

Figure 1.: Average treatment effects of group cues

Note: The figure shows treatment effect estimates with 90 and 95% confidence intervals. P -values for the null hypothesis of no differences is reported on the right.

Can individuating information mitigate discrimination?

As outlined above, half of the applications and CVs contained individuating information designed to reduce the perception of cultural distance. This section explores whether the main treatment effects are moderated by providing counter-stereotypical information. Table 3 shows an interaction model in which I regress the callback-outcome on the main treatments (alias and headscarf), the information-treatment and an interaction between the two. The interaction effects capture the degree to which the individuating information is consequential for the relative difference in callbacks between minority and majority candidates. If the individuating information reduces discrimination, the estimates should be positive. As expected, when looking at the majority applicants only, the additional information does not change the callback rate compared to the baseline applications. Surprisingly, there is no effect distinguishable from zero for the minority candidates either. For minority candidates without a headscarf, the information in fact results in a small statistically

insignificant increase in effect, whereas the opposite is true for minorities who wears a headscarf. Hence, when comparing the combined effects for the two minority candidates in column 2, it yields an interaction effect of 0.5 percentage points. In total, the evidence renders little support for the notion that adding counter-stereotypical information has a substantial effect on the relative differences in callbacks.

Table 3. Interaction models

	<i>Dependent variable:</i>	
	Callback	
	(1)	(2)
Minority	-0.031 (0.041)	
Minority + scarf	-0.138*** (0.041)	
Information	-0.018 (0.041)	-0.016 (0.041)
Minority * Information	-0.031 (0.058)	
Minority + scarf * Information	0.044 (0.058)	
Minority cues combined		-0.083** (0.035)
Minority cues combined * Information		0.005 (0.050)
Constant	0.316*** (0.029)	0.314*** (0.029)
Observations	1,350	1,350
R ²	0.013	0.008
Adjusted R ²	0.010	0.006

Note: The dependent variable is receiving a callback. *p<0.1; **p<0.05; ***p<0.01

The fact that the individuating information conveyed in the applications did not moderate the effects of the group cues is surprising. There are several possible explanations for the absence of any effects. First, discriminating employers might have sorted minority applicants based on their name and CV photographs before reading the final part of the application. In other words, some of the

discriminating employers might not have been exposed to the information treatment, making their decision based only on the explicit group cue (name and photograph). Employers receive many applications and if they use names or photographs as heuristics, they might miss the information in the applications (Bertrand and Mullainathan 2004). Second, although considerable research shows that the effects of stereotypes can be attenuated by presentation of certain types of individuating information, the effectiveness of such information depends on the amount of attention available for processing it (Andersen and Guul 2018). When subjects have limited time to form impressions – as is often the case in the hiring process – individuating information is less effective (Pratto and Bargh 1991). Finally, it is also possible that the treatment was not sufficiently powerful to reduce negative perceptions of minority applicants and cause a noticeable effect. This may especially be true given that the baseline applications were well qualified, resulting in only a modest difference between the applications with and without such information. Nonetheless, it does not ameliorate the material consequences for the individual minority applicant. The results indicate that immigrant-origin minorities can be both highly qualified and indicate an adoption of mainstream behavior, yet be targeted significant discrimination.

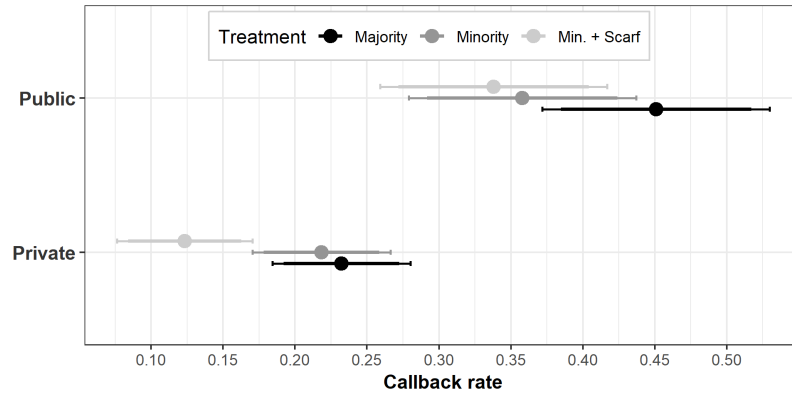
Exploratory analysis: Do results vary by sector and occupation?

This section explores how the differences in callbacks unfold across sectors and workplace characteristics.¹⁷ Figure 2 reports the differences in callback rates between the three experimental groups across sectors. Dividing the sample into subsets decreases power and inflates the confidence intervals around the point estimates. Nonetheless, several results are worth noticing. First, the hierarchy in callbacks between the treatment groups exists within both the public and the private sectors; in fact, discrimination against minorities without headscarves appears to be less common in the private sector than in the public sector. This is surprising, since all public sector job advertisements in Denmark rely on standardised hiring templates and explicitly appeal to applicants with different backgrounds. Some research has suggested that formalised hiring procedures can limit

¹⁷ Note that this section is exploratory. I did not have a strong theory leading to expectations in specific directions *a priori*, and the heterogeneous effects explored in this section were not specified in the pre-registration plan.

individual discretion and hence lessen discrimination (Reskin 2011; Midtbøen 2015). Furthermore, it has been argued that discrimination is less prevalent in public sector jobs, which deliberately address diversity and pursue compliance with anti-discrimination laws (Midtbøen 2015). Despite these organisational characteristics, discrimination was found to be rife within public sector jobs: the majority candidates received a callback rate of 45% while the minority candidates received substantively fewer callbacks (35.8% and 33.8%, respectively). This corresponds to a difference of 9.2 and 11.2 percentage points vis-à-vis the majority candidates. In private sector jobs, it is noteworthy that the difference between the majority applicants and minorities without a headscarf was only 1.4 percentage points. However, minority candidates wearing a headscarf were subject to substantive differences of 11 percentage points, implying an overall difference of 88%.

Figure 2. Callbacks across sector



Note: Callback rates for the three treatments across sectors with 90 and 95% confidence intervals.

It should be noted that the jobs applied for varied substantially across sectors. For instance, while jobs within retail are usually private sector jobs, healthcare jobs are almost exclusively public sector jobs. To differentiate between occupational categories – and to explore how specific occupations drive the average treatment effect – the callback rates are broken down by occupational categories in Appendix C. While the relative differences vary, the hierarchy is consistent: The same

pattern of differential treatment is present in all six occupational categories.

Discussion and conclusion

An enduring question concerning the integration of ethnic minorities in the labour market is what shapes interactions between employers and job seekers. The evidence from a growing experimental literature generally shows that immigrant-origin minorities become targets of discrimination even when they are highly qualified and have acquired formal skills (Vernby and Dancygier 2018; Midtbøen 2015). There is good reason to expect that perceived distance between employers and applicants in terms of cultural distinctiveness is a key component in this type of discriminatory behaviour. This further raises the question of whether individuals who have, and are able to indicate that they have, adapted to the host countries' cultural norms and way of life obtain equality in access to the labour market. In the present study, I attempt to shed light on these questions by disentangling how different explicit and implicit signals of cultural proximity, alone and in combination, affect discriminatory behaviour.

The experiment provides evidence that explicit group traits – signified by names and CV photographs – result in discriminatory behaviour by employers. Despite the high overall rate of invitations to job interviews, which illustrates the good quality of the fictitious applications and CVs, immigrant-origin applicants receive significantly fewer callbacks, especially if they also wear a headscarf. Moreover, endowing the applications with information that contradicted cultural-based stereotypes did not erase the differential treatment. In fact, it had no modifying effect and thus did not override any of the negative effects of the explicit group cues.

It is worth considering a number of limitations of the study. One concern relates to the precision of the estimates. Specifically, the difference between minority and majority candidates of 4.6 percentage points is seemingly smaller than that which is found in comparable studies that consider Middle Eastern minorities (Zschirnt and Ruedin 2016). However, the results align with recent studies that indicate that female ethnic minorities are discriminated less than male ethnic minorities (Bursell 2014; Vernby and Dancygier 2018; Dahl and Krog 2018). Moreover, the effect uncovered

in the present study is almost identical to a previous correspondence experiment conducted in the Danish labour market by Dahl and Krog (2018).

Another concern relates to the fact that the individuating information might not have been sufficiently powerful to change employers' perceptions of the applicants. In contrast to this finding, Kaas and Manger (2012) have demonstrated that adding a simple letter of reference effectively mitigates discrimination. Future work should expand this research agenda and explore which tools immigrant-origin minorities have at their disposal to escape differential treatment.

It is also worth raising three concerns related to the generalisability of the results. First, while the present study examines discrimination in the first stage of the hiring process, it remains unknown how job interviews unfold. As documented by other researchers, discrimination can occur at the stage of job interviews as well as during promotion process (Pager and Western 2012). Secondly, the present study examines discrimination in a limited sample of jobs available in the Danish labour market. While I applied for a broad variety of jobs, a number of occupational categories and job types were not included. This implies that the results may look different in other types of jobs characterised by different requirements in respect of level of education or experience. Thirdly, a general concern in field experimental studies that rely on names as proxies for group identities is that subjects are unable to infer the group cue from the name. However, considering that I used the most common names and that these names are distinct signals for non-Western immigrant-origin minorities, it is unlikely that they were misperceived by employers. Moreover, the rate of callbacks across minority names are distributed evenly which suggests that the names are perceived in a similar manner (see appendix A).

Despite these limitations, it is worthwhile considering the broader implications of the findings. The fact that discrimination persists against immigrant-origin applicants in this experiment, even when the applicants indicated strong signals of productivity-related skills *and* signals of cultural soft skills, is disconcerting. While some evidence suggests that inequalities between majorities and immigrant-origin minorities disappear once cultural differences are accounted for (Koopmans 2016), this study indicates that cues of adherence to norms do not effectively guard against discrimination. While cultural assimilation may improve immigrant-origin minorities' overall chances of finding a

job – for example by increasing their network, knowledge of the skills required and proficiency in the majority language (Statham and Tillie 2016) – they still face discrimination in the hiring process. Finally, the findings indicate that discrimination is a pervasive feature of the hiring process across sectors and job types, even if immigrant-origin minorities have assimilated to local norms. From an institutional perspective, this result grants urgency to anti-discrimination policies that can alter employers' behaviour to ensure equality in the application process.

References

- Adida, Claire L, David D Laitin, and Marie-Anne Valfort (2010). “Identifying barriers to Muslim integration in France”. In: *Proceedings of the National Academy of Sciences* 107.52, pp. 22384–22390.
- (2016). *Why Muslim integration fails in Christian-heritage societies*. Harvard University Press.
- Aggergaard, Sine (2008). “Unges idrætsdeltagelse og integration i idrætsforeninger i Århus Vest”. In: *København: Idrættens Analyseinstitut*.
- Agerström, Jens et al. (2012). “Warm and competent Hassan= cold and incompetent Eric: A harsh equation of real-life hiring discrimination”. In: *Basic and Applied Social Psychology* 34.4, pp. 359–366.
- Ahmed, Ali M (2010). “Muslim discrimination: Evidence from two lost-letter experiments”. In: *Journal of Applied Social Psychology* 40.4, pp. 888–898.
- Aigner, Dennis J and Glen G Cain (1977). “Statistical theories of discrimination in labor markets”. In: *ILR Review* 30.2, pp. 175–187.
- Andersen, Simon Calmar and Thorbjørn Sejr Guul (2018). “Reducing Minority Discrimination at the Front Line—Combined Survey and Field Experimental Evidence”. In: *Journal of Public Administration Research and Theory*.
- Andriessen, Iris et al. (2012). “Ethnic discrimination in the Dutch labor market: Its relationship with job characteristics and multiple group membership”. In: *Work and Occupations* 39.3, pp. 237–269.
- Arai, Mahmood, Moa Bursell, and Lena Nekby (2016). “The reverse gender gap in ethnic discrimination: Employer stereotypes of men and women with Arabic names”. In: *International Migration Review* 50.2, pp. 385–412.
- Baert, Stijn (2018). “Hiring discrimination: an overview of (almost) all correspondence experiments since 2005”. In: *Audit studies: Behind the scenes with theory, method, and nuance*. Springer, pp. 63–77.

- Baert, Stijn et al. (2015). “Is there less discrimination in occupations where recruitment is difficult?” In: *ILR Review* 68.3, pp. 467–500.
- Becker, Gary S (2010). *The economics of discrimination*. University of Chicago press.
- Bertrand, Marianne and Sendhil Mullainathan (2004). “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination”. In: *American economic review* 94.4, pp. 991–1013.
- Birkelund, Gunn Elisabeth, Kristian Heggebø, and Jon Rogstad (2017). “Additive or multiplicative disadvantage? The scarring effects of unemployment for ethnic minorities”. In: *European Sociological Review* 33.1, pp. 17–29.
- Blommaert, Lieselotte, Marcel Coenders, and Frank Van Tubergen (2013). “Discrimination of Arabic-named applicants in the Netherlands: An internet-based field experiment examining different phases in online recruitment procedures”. In: *Social forces* 92.3, pp. 957–982.
- Booth, Alison L, Andrew Leigh, and Elena Varganova (2012). “Does ethnic discrimination vary across minority groups? Evidence from a field experiment”. In: *Oxford Bulletin of Economics and Statistics* 74.4, pp. 547–573.
- Bowen, John Richard (2007). *Why the French don't like headscarves: Islam, the state, and public space*. Princeton University Press.
- Brader, Ted, Nicholas A Valentino, and Elizabeth Suhay (2008). “What triggers public opposition to immigration? Anxiety, group cues, and immigration threat”. In: *American Journal of Political Science* 52.4, pp. 959–978.
- Bursell, Moa (2014). “The multiple burdens of foreign-named men—evidence from a field experiment on gendered ethnic hiring discrimination in Sweden”. In: *European Sociological Review* 30.3, pp. 399–409.
- Butler, Daniel M and Margit Tavits (2017). “Does the Hijab Increase Representatives’ Perceptions of Social Distance?” In: *The Journal of Politics* 79.2, pp. 727–731.
- Byng, Michelle D (2010). “Symbolically Muslim: media, hijab, and the West”. In: *Critical Sociology* 36.1, pp. 109–129.

- Carlsson, Magnus (2010). “Experimental evidence of discrimination in the hiring of first-and second-generation immigrants”. In: *Labour* 24.3, pp. 263–278.
- Carlsson, Magnus, Luca Fumarco, and Dan-Olof Rooth (2018). “Ethnic discrimination in hiring, labour market tightness and the business cycle-evidence from field experiments”. In: *Applied Economics* 50.24, pp. 2652–2663.
- Dahl, Malte and Niels Krog (2018). “Experimental evidence of discrimination in the labour market: Intersections between ethnicity, gender, and socio-economic status”. In: *European Sociological Review* 34.4, pp. 402–417.
- DST (2018). “Indvandrere i Danmark 2018”. In: *Danmarks Statistik*.
- Economic Cooperation, Organization for and Development (2017). *Government at a Glance 2017*. OECD.
- Einstein, Katherine Levine and David M Glick (2017). “Does race affect access to government services? An experiment exploring street-level bureaucrats and access to public housing”. In: *American Journal of Political Science* 61.1, pp. 100–116.
- Ejsing, Jens (2014). *Der er langt mellem indvandrerne i foreningsdanmark*. Ed. by Berlingske.
- Erickson, Frederick and Jeffrey J Shultz (1982). *The counselor as gatekeeper: Social interaction in interviews*. Academic Press.
- Fibbi, Rosita, Mathias Lerch, and Philippe Wanner (2006). “Unemployment and discrimination against youth of immigrant origin in Switzerland: when the name makes the difference”. In: *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale* 7.3, pp. 351–366.
- Gerber, Alan S and Donald P Green (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.
- Hainmueller, Jens and Dominik Hangartner (2013). “Who gets a Swiss passport? A natural experiment in immigrant discrimination”. In: *American political science review* 107.1, pp. 159–187.
- Hainmueller, Jens and Daniel J Hopkins (2014). “Public attitudes toward immigration”. In: *Annual Review of Political Science* 17, pp. 225–249.

- (2015). “The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants”. In: *American Journal of Political Science* 59.3, pp. 529–548.
- Harell, Allison et al. (2012). “The impact of economic and cultural cues on support for immigration in Canada and the United States”. In: *Canadian Journal of Political Science/Revue canadienne de science politique* 45.3, pp. 499–530.
- Helbling, Marc (2014). “Opposing muslims and the muslim headscarf in western europe”. In: *European Sociological Review* 30.2, pp. 242–257.
- Hemker, Johannes and Anselm Rink (2017). “Multiple dimensions of bureaucratic discrimination: Evidence from German welfare offices”. In: *American Journal of Political Science* 61.4, pp. 786–803.
- Hopkins, Daniel J (2015). “The upside of accents: Language, inter-group difference, and attitudes toward immigration”. In: *British Journal of Political Science* 45.3, pp. 531–557.
- Horiuchi, Yusaku, Kosuke Imai, and Naoko Taniguchi (2007). “Designing and analyzing randomized experiments: Application to a Japanese election survey experiment”. In: *American Journal of Political Science* 51.3, pp. 669–687.
- Hughes, D Alex et al. (2017). “Continuing Evidence of Discrimination Among Local Election Officials”. In: *Available at SSRN 3089893*.
- Iyengar, Shanto et al. (2013). “Do attitudes about immigration predict willingness to admit individual immigrants? A cross-national test of the person-positivity bias”. In: *Public opinion quarterly* 77.3, pp. 641–665.
- Kaas, Leo and Christian Manger (2012). “Ethnic discrimination in Germany’s labour market: a field experiment”. In: *German economic review* 13.1, pp. 1–20.
- Konitzer, Tobias B et al. (2019). “Ethnocentrism versus group-specific stereotyping in immigration opinion: cross-national evidence on the distinctiveness of immigrant groups”. In: *Journal of Ethnic and Migration Studies*, pp. 1–24.
- Koopmans, Ruud (2016). “Does assimilation work? Sociocultural determinants of labour market participation of European Muslims”. In: *Journal of Ethnic and Migration Studies* 42.2, pp. 197–216.

- Lahey, Joanna and Ryan Beasley (2018). “Technical aspects of correspondence studies”. In: *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. Springer, pp. 81–101.
- Lancee, Bram (2019). *Ethnic discrimination in hiring: comparing groups across contexts. Results from a cross-national field experiment*.
- McGinnity, Frances and Peter D Lunn (2011). “Measuring discrimination facing ethnic minority job applicants: an Irish experiment”. In: *Work, employment and society* 25.4, pp. 693–708.
- Midtbøen, Arnfinn H (2014). “The invisible second generation? Statistical discrimination and immigrant stereotypes in employment processes in Norway”. In: *Journal of Ethnic and Migration Studies* 40.10, pp. 1657–1675.
- (2015). “The context of employment discrimination: interpreting the findings of a field experiment”. In: *The British journal of sociology* 66.1, pp. 193–214.
- (2016). “Discrimination of the second generation: Evidence from a field experiment in Norway”. In: *Journal of International Migration and Integration* 17.1, pp. 253–272.
- Ministry of Employment, MoE (2018). “Arbejdsgruppe om fremme af beskæftigelse og integration gennem deltagelse i idræt og foreningsliv”. https://kum.dk/fileadmin/KUM/Documents/Nyheder20og20Presse/Pressemeddelelser/2018/Bilag_2_rapport_fra_BM_arbejdsgruppe.pdf. unpublished.
- Norris, Pippa and Ronald F Inglehart (2012). “Muslim integration into Western cultures: Between origins and destinations”. In: *Political Studies* 60.2, pp. 228–251.
- Nunley, John M et al. (2016). “College major, internship experience, and employment opportunities: Estimates from a résumé audit”. In: *Labour Economics* 38, pp. 37–46.
- Oreopoulos, Philip (2011). “Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes”. In: *American Economic Journal: Economic Policy* 3.4, pp. 148–71.
- Oskooii, Kassra AR (2018). “Perceived Discrimination and Political Behavior”. In: *British Journal of Political Science*, pp. 1–26.
- Ostfeld, Mara (2017). “The backyard politics of attitudes toward immigration”. In: *Political Psychology* 38.1, pp. 21–37.

- Pager, Devah and Bruce Western (2012). "Identifying discrimination at work: The use of field experiments". In: *Journal of Social Issues* 68.2, pp. 221–237.
- Phelps, Edmund S (1972). "The statistical theory of racism and sexism". In: *The american economic review*, pp. 659–661.
- Pierné, Guillaume (2013). "Hiring discrimination based on national origin and religious closeness: results from a field experiment in the Paris area". In: *IZA Journal of Labor Economics* 2.1, p. 4.
- Pratto, Felicia and John A Bargh (1991). "Stereotyping based on apparently individuating information: Trait and global components of sex stereotypes under attention overload". In: *Journal of Experimental Social Psychology* 27.1, pp. 26–47.
- Quillian, Lincoln et al. (2017). "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time". In: *Proceedings of the National Academy of Sciences* 114.41, pp. 10870–10875.
- Reskin, Barbara F (2011). "Rethinking employment discrimination and its remedies". In: *The inequality reader: Contemporary and foundational readings in race, class, and gender*, pp. 378–88.
- Riach, Peter A and Judith Rich (2004). "Deceptive field experiments of discrimination: are they ethical?" In: *Kyklos* 57.3, pp. 457–470.
- Rivera, Lauren A (2012). "Hiring as cultural matching: The case of elite professional service firms". In: *American sociological review* 77.6, pp. 999–1022.
- Schildkraut, Deborah J (2010). *Americanism in the twenty-first century: Public opinion in the age of immigration*. Cambridge University Press.
- Shadid, Wasif and P Sjoerd Van Koningsveld (2005). "Muslim dress in Europe: debates on the headscarf". In: *Journal of Islamic Studies* 16.1, pp. 35–61.
- Sides, John and Jack Citrin (2007). "European opinion about immigration: The role of identities, interests and information". In: *British journal of political science* 37.3, pp. 477–504.
- Simonsen, Kristina Bakkær (2017). "Integration af indvandrere i Danmark." In: *Politica: Tidsskrift for Politisk Videnskab* 49.3.

- Slot, Line Vikkelsø (2008). *Praktikpladser og formidlingspraksis: praktikkonsulent i spændingsfeltet mellem virksomhedskrav og hensynet til ikke-diskrimination*. Rockwool Fondens Forskningsenhed.
- Sniderman, Paul M, Louk Hagendoorn, and Markus Prior (2004). “Predisposing factors and situational triggers: Exclusionary reactions to immigrant minorities”. In: *American political science review* 98.1, pp. 35–49.
- Statham, Paul and Jean Tillie (2016). “Muslims in their European societies of settlement: a comparative agenda for empirical research on socio-cultural integration across countries and groups”. In: *Journal of Ethnic and Migration Studies* 42.2, pp. 177–196.
- Vernby, Kåre and Rafaela M Dancygier (2018). *Employer discrimination and the immutability of ethnic hierarchies: A field experiment*. Tech. rep. Working Paper.
- Villadsen, Anders R and Jesper N Wulff (2018). “Is the Public Sector a Fairer Employer? Ethnic employment discrimination in the public and private sectors”. In: *Academy of Management Discoveries* 4.4, pp. 429–448.
- Vuolo, Mike, Christopher Uggen, and Sarah Lageson (2018). “To match or not to match? Statistical and substantive considerations in audit design and analysis”. In: *Audit studies: Behind the scenes with theory, method, and nuance*. Springer, pp. 119–140.
- Weichselbaumer, Doris (2016). “Discrimination against female migrants wearing headscarves”. In: Wright, Bradley RE et al. (2013). “Religious affiliation and hiring discrimination in New England: A field experiment”. In: *Research in Social Stratification and Mobility* 34, pp. 111–126.
- Zschirnt, Eva (2016). “Revisiting ethics in correspondence testing”. In: *NCCR-on the move, working paper series* 8.
- Zschirnt, Eva and Didier Ruedin (2016). “Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015”. In: *Journal of Ethnic and Migration Studies* 42.7, pp. 1115–1134.

Supplementary material for "Alike but different: How cultural distinctiveness shapes immigrant-origin minorities' access to the labour market"

Appendix A. Treatment details

This appendix shows details on the treatment. Table A1 gives a list of the aliases used as indicators for ethnicity. In total, the possible combinations of majority names are 96, while the possible combinations for minority aliases are 64. The names were chosen from a data set comprising the number of Danish citizens with a given first or last name. The immigrant-origin minority first names all figure in a dictionary of Middle Eastern first names that are in frequent use in Denmark (Meldgaard, 2005), and were identified from among the 20 most popular Middle Eastern names in use in Denmark. Half of the popular ethnic majority first names were among the top 20 most popular female names. Following Dahl and Krog (2018), half of the majority names are names with low-socioeconomic connotations in order to explore if socioeconomic perceptions affect callback rates. This was not the case. The low-SES names are based on a population index following Dahl and Krog (2018).¹⁸ The selected low-SES names all belong to individuals in the bottom decile of the SES index with an average age between 20 and 40. Finally, the last names were chosen from among the most popular Danish-sounding and Middle Eastern-sounding last names.

¹⁸ The SES-index was constructed using three items: average crime rate, unemployment rate, and annual income. The crime rate is a measure of the proportion of the population with a given name who have been incarcerated within the last five years. The unemployment statistic is a measure of the proportion of people with a given name who receive unemployment benefits or have been unemployed for at least six months. Income is the average annual income earned by those bearing a given name. All items were scored from 1 to 8, where a higher score indicates lower status (with income scored in reverse order).

Table A. Information in job applications

Majority first names	Majority last names	Minority first names	Minority last names
Anna (0.38)	Andersen (0.37)	Aisha (0.28)	Abdi (0.14)
Anne (0.42)	Christensen (0.30)	Amal (0.18)	Farah (0.15)
Belinda (0.32)	Hansen (0.38)	Amira (0.22)	Ismail (0.25)
Bonnie (0.34)	Jensen (0.32)	Fatima (0.20)	Mohamed (0.26)
Hanne (0.31)	Larsen (0.28)	Fatma (0.20)	Osman (0.20)
Helle (0.26)	Nielsen (0.29)	Hatice (0.22)	Sahin (26)
Jennie (0.29)	Pedersen (0.22)	Iman (0.22)	Yildiz (0.28)
Jennifer (0.40)	Sørensen (0.30)	Zainab (0.28)	Yilmaz (0.24)
Maria (0.25)			
Mette (0.26)			
Michelle (0.17)			
Stella (0.17)			
p : 0.32	p : 0.72	p : 0.64	p : 0.21

Note: Callback rates are reported in parentheses. Final line shows the p-value from F-tests for the hypothesis that callbacks to candidates across names are equivalent.

Figure A. CVs showing applicants with and without headscarf

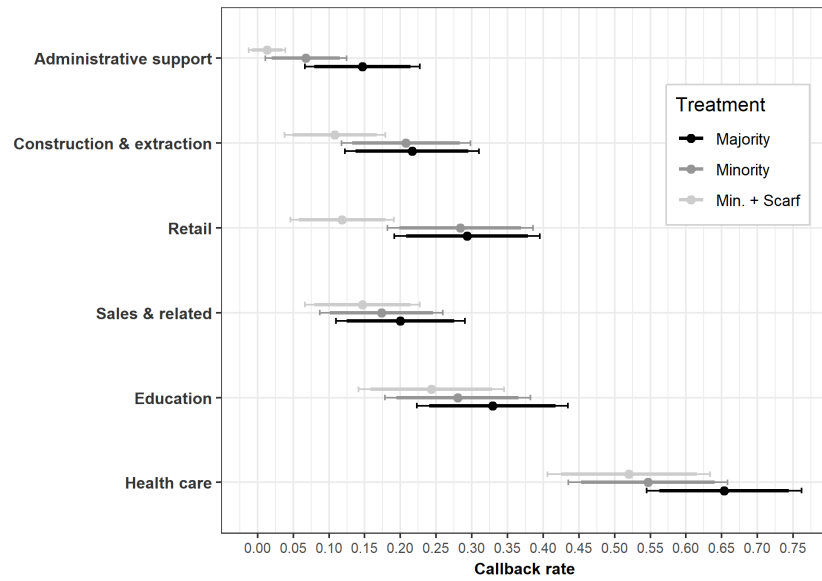
Note: Example of CVs translated from Danish.

Appendix B. Ethical considerations

There is a large literature of the ethical considerations involved when conducting correspondence experiments including advice on how to minimize harm to subjects involved (see Zschirnt (2016) and Riach and Rich (2004) for an extensive discussion of ethics in correspondence experiments). Central arguments for the legitimacy of carrying out correspondence experiments include that (i) the question of discrimination is of high societal importance, (ii) that there is no other way to credibly retrieve this vital information, and (iii) if the research is prepared and carried out carefully, there is only very limited detrimental effect on the employers tested. On the last point, this study partly replicates a previous experiment that had confirmed that the experimental setup and the logistic of sending and handling the callbacks operated efficiently. Further, in order to ensure sufficient statistical power without having to contact an excessively large number of employers, I conducted power analysis before the collection of data. Finally, I carefully considered how to minimize any inconveniences to the employers that were contacted as a part of the experiment. The main cost to employers is time, so all requests were answered as quickly as possible, explaining that the applicant had just found another job. Data is analyzed on an aggregated form to keep all individual employers anonymous in order to avoid associations between specific companies and this study.

Appendix C. Discrimination across sectors

Figure C. Callbacks across occupational categories



Note: The figure shows callback rates for the three treatments across occupational categories with 90 and 95% confidence intervals.

Who is responsive? How electoral incentives and candidate selection shape ethnocentric responsiveness

Malte Dahl, Peter T. Dinesen and Mikkel Schioeler*

Presented at the Midwest Political Science Association Conference, April 2019

Abstract

Previous studies have documented racial and ethnic biases in politicians' constituency service. We propose and test two mechanisms for mitigating such ethnocentric responsiveness. First, we ask whether sufficiently strong electoral incentives motivate incumbent politicians to serve their constituents and crowd out potential biases (the electoral competitiveness mechanism). Second, we examine whether minority constituents are able to identify responsive legislators by using candidates' partisan affiliation and their stated policy preferences as heuristics (the candidate selection mechanism). To do so, we conducted a field experiment prior to the Danish local elections in 2017 in which local incumbents received a simple request from constituents with randomly assigned descriptive characteristics (ethnicity and gender) and intentions to cast a personal vote ($N = 2,395$). Moreover, we collected data on electoral performances and use a voting advice application to gauge candidates' stated policy preferences. We find that politicians are substantially less likely to respond to constituents who do not share their ethnicity. Although the overall responsiveness increases with stronger incentives, there is no support for the notion that electoral incentives mitigate biased behavior. However, the results indicate that minority constituents can effectively use parties' positions and individual candidates' policy preferences concerning immigration issues as a means to identify responsive politicians. The results shed new light on the consequences of biases in legislator responsiveness and hold important institutional consequences.

*We would like to thank Jens Olav Dahlgaard, David Broockman, Lene H. Pedersen, Benjamin Egerod, Frederik Hjorth, participants at MPSA 2019. Department of Political Science, University of Copenhagen, 1353 Copenhagen.

Introduction

Political responsiveness and political equality are fundamental aspects of representative democracy. In theory, a representative democracy with universal suffrage secures interest representation of various social groups by enabling them to elect their preferred legislators. Yet, this is often not the case in practice (Mansbridge 1999; Grose 2011; Canon and Posner 1999). Understanding the nature of such underrepresentation and, concomitantly, what, if anything, that can be done to obtain parity in representation, are core questions for political science as their answers carry implications for fundamental discussions about the quality and inclusiveness of democracy.

A large body of research demonstrates that politicians are more likely to advance the preferences of constituents with whom they share personal characteristics (Burden 2007; Carnes 2012; Nye, Rainer, and Stratmann 2014; Iyer et al. 2012). For example, white legislators are significantly less likely than black legislators to support policies that specifically benefit black communities (Grose 2011), and race and ethnicity influence how much time and effort politicians spend on behalf of their constituents in oversight of bureaucratic policy-making (Minta 2009). There is also ample evidence that such in-group favoritism manifests itself directly in the legislator-constituent relationship. Several recent field experimental audit studies of legislator responsiveness to simple constituent requests – a simple, but face valid indicator of legislator attention to constituent concerns – show that politicians are more responsive to citizens with whom they share race or ethnicity (Butler and Broockman 2011; Broockman 2013; Gell-Redman et al. 2018; Mendez and Grose 2018). For example, when state legislators are asked for help with voter registration, black voters are significantly less likely to receive an answer from white legislators, whereas the reverse pattern is true for black legislators responding to white constituents (Butler and Broockman 2011; McClendon 2016). Because racial and ethnic minority groups tend to be descriptively underrepresented among elected politicians (Bloemraad 2013; Warshaw 2019), such ethnocentric¹ legislator behavior implies unequal access to policy-making – and therefore unequal inclusion in democracy – for these minority

¹ Researchers have identified distinct motives for such bias including strategic motives, out-group prejudice, in-group favoritism, and linked fate (Butler 2014). While these are distinct concepts, they hold similar consequences. We simply use the term ‘ethnocentrism’ or ‘in-group favoritism’ to denote preferential treatment of in-group members.

groups.

Departing from the pervasive finding of ethnocentric responsiveness,² and existing work on the underlying motivations that animate such legislator behavior, this paper asks how, if at all, such tendencies can be curbed? That is, under which circumstances do parity in responsiveness occur? Ultimately, this is the question of interest from the perspective of ethnic minorities, who experience lower legislator responsiveness and poorer constituency service more generally. More specifically, we contribute to the debate about legislator responsiveness by scrutinizing two different mechanisms that may plausibly further equal responsiveness. Originating in pluralist theories of democracy (Dahl 1967), the first mechanism – which we will call *electoral competitiveness* – builds on the assumption that politicians are extrinsically motivated, specifically in regard to obtaining (re)election, and therefore respond to electoral incentives (Alt, Bueno de Mesquita, and Rose 2011; Fenno 1978). In essence, this “structural” perspective predicts that if elections are sufficiently competitive, this will crowd out other concerns than re-election, and in turn animate instrumental legislators to be responsive to their constituents, independent of their ethnic background (or any other characteristic), in order to gain their votes (Fenno 1978).

The second mechanism, which we refer to as *candidate selection*, makes no assumptions regarding the motivations of legislators, but instead puts the individual voter in focus, by emphasizing her or his ability to identify politicians that are (more) responsive to them. Because acquiring the information necessary to determine responsive politicians is demanding, we expect voters to resort to heuristics – cognitive shortcuts – that reduce these costs (Lau and Redlawsk 2001; Lupia, McCubbins, Arthur, et al. 1998). Specifically, we examine two types of heuristics regarding legislator responsiveness: their partisan affiliation and their stated policy preferences. The party serves as a “low-cost” heuristic, potentially enabling voters to identify groups of responsive politicians on average. Obtaining individual legislators’ stated preferences is more costly in information seeking terms, yet potentially more efficient as politicians may vary in responsiveness *within* parties. The

² Researchers have identified a range of explanations for such bias including pursuit of strategic motives, out-group prejudice, in-group favoritism, and a linked fate within groups (Butler 2014). While these are distinct concepts, they hold similar consequences. Therefore, we simply use the terms ‘ethnocentrism’ and ‘in-group favoritism’ to denote preferential treatment of in-group members.

key issue here is to identify the relevant policy preferences by which to select legislators.

We assess the proposed mechanisms for securing legislator responsiveness to immigrant-origin minorities by combining a field experiment eliciting the responsiveness of incumbent local politicians in Denmark, asking for their help in locating polling stations in the upcoming municipal election, with auxiliary observational data on electoral performance (seat winning margin), politicians' decision to run for re-election, as well as their stated preferences on a range of policy areas gauged via a voting advice application. Following previous work, in the request, each politician was randomly treated with a name ("alias") that signaled *ethnicity* (majority Danish/non-western immigrant-origin), *gender* (male/female), as well as *personal voting intention*, specifically stating the intention to vote for the politician in question in the upcoming election. The observational data serve to assess the two mechanisms in the field experimental data through analyses of heterogeneous effects.

We report several important findings. First, as a baseline result, we replicate previous findings of unequal responsiveness – from the US and South Africa – in the Danish setting, thereby indicating that in-group favoritism is a more general phenomenon and not confined to specific racial or ethnic out-group constellations. Second, focusing on majority legislators, we address the electoral competitiveness mechanism by analyzing responsiveness among a subset of legislators with strong electoral incentives for responding to constituent requests. Specifically, legislators who decided to (i) seek reelection, (ii) who won their seat by a narrow margin, and, most importantly, (iii) who received a request stipulating the voters' intention to vote for them. We find little evidence that any of these incentives – alone or in combination – influences majority legislators' ethnocentric responsiveness. Third, we scrutinize the candidate selection mechanism, and the predicted use of two heuristics, by breaking down responsiveness by, first, legislator's party affiliation and, second, their individually stated policy preferences vis-à-vis immigration – the political issue we expect to be most predictive of ethnocentric responsiveness. We find that using party affiliation is an effective strategy for locating legislators that are more responsive; in fact, the most pro-immigration parties are equally responsive to ethnic majorities and minorities. However, among the remaining parties, including the three largest centrist parties, ethnocentric responsiveness is observed. Yet,

within parties, there is considerable variation in immigration policy preferences between individual candidates. We find that these preferences are strongly predictive of responsiveness within parties, thus implying that even within a relatively immigration-sceptic party (e.g., the centre-right party *The Liberal Party* or the center-left party, *The Social Democratic Party*), higher responsiveness can be obtained by locating a pro-immigration candidate.

Group-centric responsiveness: previous work and our approach

Responsiveness to constituency concerns and questions expressed in their queries is a principal activity of local politicians and an essential feature of legitimate democratic representation (Grose 2011; Fenno 1978). Assisting constituents, especially with answering service questions unrelated to policy issues, is also a relatively easy way for politicians to reach out to voters to gain trust and enhance their reputations (Fiorina 1989; Butler, Karpowitz, and Pope 2012; Cain, Ferejohn, and Fiorina 1987). Thus, high-quality and prompt responsiveness offers a pathway for politicians to maximize their electoral prospects. This is especially true in local elections where a few personal votes are often decisive for the electoral outcome.

Despite the potential electoral benefits of being responsive to their constituents, existing research demonstrates that politicians exhibit in-group favoritism in their communication to constituents. Specifically, several recent field experiments on legislator responsiveness – predominantly in the US context – show that politicians are more likely to respond to citizen requests for constituency service from those with whom they share a racial or ethnic background (e.g. Butler and Broockman (2011), Broockman (2013), Gell-Redman et al. (2018), and Mendez and Grose (2018)).³ For example, when state legislators are asked for help with voter registration, black voters are significantly less likely to receive an answer from white legislators, whereas the reverse pattern is true for black legislators responding to white constituents (Butler and Broockman 2011). Similarly, American legislators are less responsive to Latino and Asian constituents compared to white constituents (Gell-Redman

³ A related literature focusing on bureaucrats (i.e. policy implementation) suggests that immigrant-origin minority citizens are also less likely to receive a reply when contacting local election registrars for voter ID requirements (White, Nathan, and Faller 2015; Hughes et al. 2017).

et al. 2018; Mendez and Grose 2018; Mendez 2014). Considering that these groups are politically underrepresented relative to ethnic or racial majorities, the differential treatment of legislators implies that they are disadvantaged in their access to policy-making. Moreover, the lack of quality and inclusiveness of political representation negatively affects minorities' political participation and trust in government (Mansbridge 1999; Banducci, Donovan, and Karp 2004).

The observed unequal responsiveness has raised the question of what motivates legislators to be more or less responsive to ethnic in- and out-groups. Theoretically, the observed ethnocentric responsiveness has primarily been viewed through the lens of legislators' underlying motivations, distinguishing between strategic (instrumental) considerations, typically related to (re-)election (*extrinsic motivations*) and non-instrumental considerations relating to values (e.g. out-group distaste or in-group loyalty) (*intrinsic motivations*).

According to the former perspective, utility maximizing politicians respond to constituent requests, including non-policy ones, as a way to vote-maximize (Fenno 1978). It follows from this proposition that politicians act in a strategic manner by investing fewer resources in voters less decisive for their reelection chances. Since constituents' personal characteristics might serve as a cue of partisanship and the propensity to participate and vote in elections, such motives may explain why political candidates – especially within parties not widely supported by racial or ethnic minorities – are less responsive to minority voters (Broockman 2013; Grose 2011).

The other class of motives – intrinsic motivations – suggests that legislators are animated by a private preference for others' wellbeing' (Ariely, Bracha, and Meier 2009, p. 544). However, this often manifests itself asymmetrically so that they advance the interests of voters that share their own personal characteristics due to in-group affirmation or loyalty (Broockman 2013). The underlying motivations for legislators' lack of responsiveness toward racial and ethnic minority groups are difficult to elucidate and disentangle, but existing evidence indicates that while strategic incentives matter (Janusz and Lajevardi 2016; Gell-Redman et al. 2018), this mechanism explains only a portion of the observed differential treatment (Butler and Broockman 2011). In the absence of any extrinsic motivations, legislators still favor in-group constituents (Broockman 2013).

Two mechanisms for obtaining equal responsiveness

In this paper, we build on the existing research on ethnocentric responsiveness and the underlying motivations outlined above and ask a related, but different research question than previous studies. Specifically, departing from the prevalent finding of ethnocentric responsiveness, we explore whether and how this differential treatment can be mitigated. Pinpointing the circumstances under which parity in responsiveness occurs is an important first step in grasping the options for countering it.

We scrutinize two different mechanisms for obtaining equal responsiveness among legislators: electoral competitiveness and candidate selection. The first mechanism, electoral competitiveness, departs from the above-mentioned assumption that politicians are motivated by strategic considerations and therefore seek office by investing more time appealing to voters who are likely to cast a vote for them (Fenno 1978; Mayhew 1974). As a consequence, if politicians compete in sufficiently competitive elections we would expect them to be more responsive to their constituents independent of their ethnic background because they cannot afford to be irresponsible to certain parts of the electorate. As such, the electoral competitiveness mechanism is “structural” in the sense that it emphasizes an aspect of the electoral system as a means to override potential ethnocentrism in responsiveness. Based on this rationale, we hypothesize that an increase in electoral competitiveness will result in (i) higher overall responsiveness, and, most importantly for our research question, (ii) less differential treatment based on ethnicity. This notion is supported by previous studies suggesting that electorally unsafe legislators are more responsive to their constituents (Cain, Ferejohn, and Fiorina 1987; Freeman and Richardson Jr 1996; Ashworth and Mesquita 2006). To test the electoral competitiveness mechanism for curbing responsiveness bias, we explore how majority legislators’ responsiveness varies by increasingly strong electoral incentives, including an experimental manipulation of a personal vote cue. Because most parties in the studied local elections rely on an open-list system, where each vote matters and a handful of votes often determine electoral outcomes, this is an optimal setting for exploring the link between competitiveness and ethnocentric responsiveness.

The second mechanism, *candidate selection*, switches focus to the individual voter and her or his ability to identify politicians who are likely to be responsive. It is generally challenging to identify such legislators, which makes it likely that voters use complexity-reducing heuristics – cognitive

shortcuts – to select a candidate (Sniderman, Brody, and Tetlock 1993; Popkin and Popkin 1994). More specifically, we expect two heuristics to be particularly effective in this regard. First, as a simple, but likely effective, heuristic, immigrant-origin constituents may be able to use parties’ policy images on specific issues as a signal of the likelihood of responsiveness to requests for help. Specifically for minority constituents, parties with more immigration-friendly policy preferences may be less susceptible to exhibiting a bias towards immigrant-origin minority constituents. Indeed, the ability of a party-label to serve as a heuristic for constituents to identify responsive legislators is supported by evidence from the US context, where studies have demonstrated a partisan divide in legislators’ bias towards Latino constituents (Gell-Redman et al. 2018; Janusz and Lajevardi 2016). The Danish multiparty system provides a good case for advancing the theory that parties’ positions on issues concerning minority groups also shape interactions with minority constituents. Because of the large number of parties, we can leverage greater variation in party positions related to questions of immigration and integration, from parties running on highly restrictionist immigration policy platforms to parties that represent very liberal ones.

Second, in addition to relying on party heuristics, constituents may resort to a higher “information cost” strategy in terms of using stated legislator preferences on specific policy issues (specifically relating to minorities), to identify responsive legislators. Mendez and Grose (2018) demonstrate that legislators who vote for restrictive voting laws are more likely to exhibit responsiveness bias towards Latino constituents. This suggests that candidates’ individual policy preferences are indicative of their responsiveness to ethnic minority constituents.⁴ In our case, the equivalent policy area is that related to immigration and integration of immigrants. This is especially important since, as we show below, incumbents’ positions on questions of immigration are not necessarily homogeneous within parties.

This *candidate selection* mechanism sheds light on the implications of responsiveness bias observed in previous work. If the differential treatment spans parties and extends to legislators who explicitly voice policies that favor minority groups, it leaves immigrant-origin constituents with few

⁴ A few studies suggest that racial majority legislators who explicitly voice the interests of minority and immigrant-origin populations can in fact engage in greater substantive representation than those of minority backgrounds (Browning, Marshall, and Tabb 1984; Siemiatycki 2009).

tools to reach equality in responsiveness. To measure stated preferences, we utilize the fact that, prior to the experiment, a majority share of the incumbents seeking re-election participated in a publicly available voting advice application, *The Candidate Test*, intended to inform voters about the positions of their local candidates on key political topics. This data enables us to examine individual candidates' stated policy preferences on immigration issues.

Research design and data

To determine whether local incumbents are responsive to simple requests from their constituents, we conducted a field experimental audit study in the context of the 2017 local elections for municipality councils in Denmark. The experiment involved sending a request to all incumbent politicians with a publicly available email address approximately six weeks prior to the elections. To answer our research questions, we merge data on the incumbents from several sources to the field experimental data. These include a host of personal characteristics (party affiliation, decision to stand for re-election, their election outcomes, and stated policy preferences on immigration), as well as characteristics of their municipality.

Empirical context

There are 98 municipalities in Denmark, which range between 2,000 and 600,000 inhabitants. The members of the city councils are elected every four years under a proportional electoral system. For the 2017 election, 9,558 candidates were competing for 2,432 seats in the local councils. The legislators receive a substantial salary (varying by committee membership etc.), but typically only mayors are full-time politicians.

A number of features of Danish municipalities make them a good setting for studying potential bias in responsiveness among legislators. Firstly, municipalities are politically and administratively highly significant in Denmark. The municipality councils are responsible for around 50% of total government expenditure, including services such as schools, infrastructure and social benefits, and they enjoy a large degree of financial and political autonomy. Local elections thus have real con-

sequences for constituents, which in turn means that they enjoy high attention from parties, the media, and voters. High turnout – 70.8% in the last election – is one clear indication of that. Secondly, important for our research questions regarding ethnocentric responsiveness, foreign citizens are allowed to vote provided they hold a residency permit and have lived in Denmark for more than three years prior to the election. In other words, even individuals with a status as refugees can vote after three years residency. This makes requests to legislators from immigrants a likely scenario. Thirdly, with a few exceptions – most notably the far left party (the Unitarian List) – the established parties use an open list-system in most municipalities. Hence, personal votes within the party decide which candidates win a seat. Further, the median candidate across all major parties received less than 100 votes, thereby indicating that for many candidates, a few additional votes can be decisive for the chances of winning a seat. By implication, they have a strong electoral incentive for constituency service in general. Fourth and lastly, most local politicians in Denmark report that they use requests from citizens to impact agendas on the councils (Dahlggaard et al. 2009; Pedersen et al. 2013). There is thus strong reason to believe that local legislators take such requests seriously, and therefore potentially also respond to them.

Designing requests

Incumbents all received one email from a voter six weeks prior to the municipal election. The requests stated that the constituent had recently moved to a new address within the same municipality and therefore was unsure about the geographic location of the polling station. The choice of this specific request reflects five considerations. First, it presented incumbents with an opportunity to provide critical constituency service. Second, such requests are common to incumbent local politicians, thus giving the treatment high ecological validity. Third, the question is easy to answer since information on where to vote is provided on the voting ballot that is mailed to all voters prior to the election. Fourth, this also implies that answering the request kept the burden imposed on legislators' time at a minimum, which is also reflected in the actual replies with a median answer length of 30 words (see appendix F for further ethical considerations). Fifth and finally, the request allowed us to include, in an organic way, a randomly assigned explicit voting preference – to

strengthen electoral incentives – by stating that the voter intended (as in the last election) to cast a personal vote for the candidate.

Treatments and randomization

In order to examine the effect of ethnic in- and out-group cues and electoral incentives, we randomly assigned incumbents to configurations of three treatments (2x2x2): the requesters' ethnic affiliation and gender, and whether the requests included an explicit vote preference. Ethnicity and gender are manipulated through names ("aliases") holding distinct Danish (majority) or Non-Western (minority) as well as male or female connotations.⁵ We relied on a combination of the five most popular traditional majority/native Danish female and male first and last names as well as the five most popular Non-western first and last names (Statistics Denmark, 2015).⁶

Finally, our manipulation of a personal voting preference for the politicians was included in half of the requests by the statement 'I intend to cast my vote for you again.' This personal vote cue is a stronger incentive than what has been used in previous research by randomizing party membership (Butler and Broockman 2011; McClendon 2016) or simply an intention to vote (Gell-Redman et al. 2018). Moreover, the strength of the personal voting cue used in the present experiment is emphasized by the fact that the local elections are often determined with small vote margins (Dahlggaard 2016). The voter request is reproduced in Figure 1 below. We randomly assigned incumbents to treatment groups using block randomization by five variables that we believed to be predictive of the outcome variable: the size of the municipality, the share of ethnic minority inhabitants, incumbents' political leaning (left or right), gender, and ethnicity.⁷

⁵ We used Gmail-accounts with email-addresses consisting of the treatment name plus a random number (see details in appendix A).

⁶ One concern is that the names used to signal ethnic affiliation serve as imprecise proxies of ethnicity. By regressing the outcome on the various minority aliases, we show that there are no significant differences in response rates across names, which indicates that specific minority names are not mistakenly perceived as proxies for majority names (Appendix A).

⁷ The incumbents' ethnic affiliation was classified by two coders as either majority or non-western based on their surnames and first names. The block randomization was conducted in R using the package *randomizR* (Coppock 2016).

Figure 1. Voter request

<p>Subject line: Polling station</p> <hr/> <p>My name is [name]</p> <p><i>[I intend to vote for you once again at the upcoming Municipal election]</i></p> <p>Since the last election, I have changed my address, and now I am unsure whether I shall use a different polling station. Can you please tell me how to retrieve this information?</p> <p>Thank you very much</p> <p>[name]</p>

Implementation details

Before launching the experiment, we conducted a pilot study ($N = 50$) to test our experimental protocol. The initial request included in the pilot was formulated slightly differently, and from the replies it was clear that a few incumbents misunderstood the question, which led us to modify the wording of the request to avoid such misunderstandings.⁸ The pilot data were not included in the final sample. The requests were emailed with five-minute intervals in batches of 60 requests at a time. This procedure enabled us to send all emails at the same evening between 6 pm and 10 pm, thus minimizing extraneous variation from events etc. between requests. Moreover, the time interval between the requests reduced the risk that two or more incumbents in the same location (e.g. at a campaign meeting) would receive an identical message at the same time, which could raise suspicion and thus imply interference between subjects. We gathered responses until the voting booths closed on the day of the election, thereby giving incumbents 42 days to respond. 90% of the replies arrived within a day and 96% within three days.

⁸ The initial question was: ‘Where should I cast my vote’. A few subjects in the pilot believed the question concerned which party to vote for rather than the geographical position of the polling station.

Measures

Outcome variables

Our main dependent variable is a dichotomous response variable (response to request vs. non-response). Since we technically treat incumbents' email addresses, not the incumbents themselves, some replies could potentially have been sent from incumbents' staff members. We observed only 16 instances where legislators were not personally responsible for answering to the requests and classifying the answers as non-replies has no effect for the analysis. We also received 12 automated responses, which we chose to code as non-responses in the analysis due to their impersonal and non-helpful nature (Costa 2017a). Other studies have documented that by focusing exclusively on whether political elites or public officials respond or not, and not how they respond, one may miss important qualitative nuances (Hemker and Rink 2017; White, Nathan, and Faller 2015). We therefore complement the dichotomous outcome with a measure that also takes the content of each individual email into account using a qualitative measure inspired by the approach described by (Costa 2017b). In Appendix C, we replicate the baseline results using the qualitative measures as outcomes.

Measuring electoral incentives

As noted above, we randomly assign a personal voting cue by which half of the requests included the statement "I intend to cast my vote for you again." In addition to the experimental manipulation of the personal voting intention, we use two observational measures indexing electoral incentives to gauge legislators' extrinsic motivation for constituency service. Parallel to earlier studies, the first is a simple dichotomous indicator of whether incumbents are re-running for election, in which case they should be more inclined to respond compared to incumbents in their final term (Butler and Broockman 2011; Butler 2014; Butler, Karpowitz, and Pope 2012). The second measures closeness of the elections, which is a proxy for incumbents' perceived likelihood of maintaining their seat. Some incumbents have safe seats, winning by a wide margin (clear winners), while others face a more competitive election over marginal seats. Incumbents faced with a risk of not being

re-elected arguably have a stronger incentive to respond to requests compared to clear winners. This is premised on local politicians having a good idea about their chances of being elected before the actual election.⁹ To measure individual candidates' electoral competition in the Danish PR-system, we used a bootstrapping-method inspired by Kotakorpi, Poutvaara, and Terviö (2017) and Dahlgaard (2016). This procedure measures all candidates' share of re-elections in 10,000 alternative elections by resampling from the actual distribution of votes (See details in Appendix C).

Since neither running for re-election and electoral competitiveness is randomly assigned, this contextual variation is not exogenous vis-à-vis responsiveness. However, in combination with the experimentally assigned personal vote cue, it is possible to identify candidates, who have very strong reasons to answer depending on whether they ran for reelection, faced close elections and received a request stating a personal vote intention. We believe this provides a rigorous test of whether electoral incentives potentially moderate in-group favoritism.

Measuring stated policy preferences

To measure parties' and incumbents' policy preferences, we used a questionnaire (voting advice application) fielded by a major Danish news outlet prior to the experiment in which individual candidates report their stances on 15 policy questions relevant for municipal politics. The answers are publicly available in an online template such that voters are able to answer the same questions as the politicians to find out which candidates best represent their opinions. As noted, we focus specifically on stated immigration policy preferences as a heuristic for responsiveness for ethnic minorities (although we also incorporate other preferences in auxiliary analyses). We measure these preferences by two available questions: 'Local municipal institutions try too hard to accommodate to religious minorities' (reversed) and 'The municipality should accept more refugees.' Both questions are answered on a 5-point scale (for descriptive statistics, see Appendix D). The measure takes on values from 0 (most negative) to 8 (most positive), with a mean of 4.3 (SD=2.4). 1,514 majority

⁹ Incumbent politicians can for example base their electoral chances on their result in the last election, their party's popularity in the municipality, how their party fares in national polls, their place on the voting ballot, and from speaking to voters and taking parting in campaign activities. Recall that the experiment was conducted only six weeks prior to the election.

incumbents answered the questionnaire just before the experiment was fielded.

Results

In total, we received 1,504 replies to the 2,395 requests sent by email, corresponding to an overall response rate of 62.8 percent. This is a high response rate compared to results from previous studies on political responsiveness (Costa 2017a), thus indicating relatively high political responsiveness in general among local politicians in Denmark. Yet, our interest is in the potential variation in responsiveness toward individuals with ethnic minority background.

Experimental baseline results: Ethnocentric political responsiveness in Denmark

As a first-order question, we examine the effect of our treatment of primary interest – the ethnic alias of the sender – as well as the two other treatments (gender of the sender and the explicit personal voting intention) on the response rate. The gender cue had no effect on the likelihood of receiving a reply neither in general nor across incumbents’ own gender or interacted with the other treatments (see Appendix E). This is an interesting finding in its own right and aligns with results from the US context (Butler 2014), but it is outside the focus of our paper. Hence, for the remainder of the paper we focus on the ethnic alias- and the vote intention-treatment.

Table 1 reports the average treatment effect of the ethnic minority alias (column 1) as well as the interaction with incumbent politicians’ ethnic background (column 2). The results show that requests signed by an ethnic minority alias received 16.2 percentage points fewer replies compared to those sent by an ethnic majority alias. The differential treatment also materializes in responses that are on average less friendly, less accurate, less timely and less elaborate (see appendix B). By interacting the ethnic alias with legislators’ own ethnic background, we can explore if the observed irresponsiveness is primarily a function of majority aversion towards the minority or ethnic in-group favoritism more generally. The results reported in column 2 and visualized in Figure 2 strongly indicate a general ethnocentrism in responsiveness. Majority incumbents are 17.2 percentage points ($N = 2,326$; $p < 0.01$) more likely to reply to requests from an ethnic in-group member compared

to requests from an out-group member. However, strikingly parallel to this, minority incumbents are 19.9 percentage points ($N = 69$; $p = .08$) more likely to respond to requests from ethnic in-group members compared to requests from out-group members (although, note the small number of observations and associated statistical uncertainty). The strong pattern of ethnocentrism in responsiveness among legislators in Danish local councils confirms findings from the US showing that incumbents in general provide better service to ethnic in-group constituents. If anything, the effects are stronger in the Danish context.¹⁰ Ethnocentric responsiveness is thus not confined to the US context with its' strong tendency for ethnic (racial) voting, but instead appears to be a phenomenon extending to settings with much less ethnicized politics.

Table 1. OLS regression results

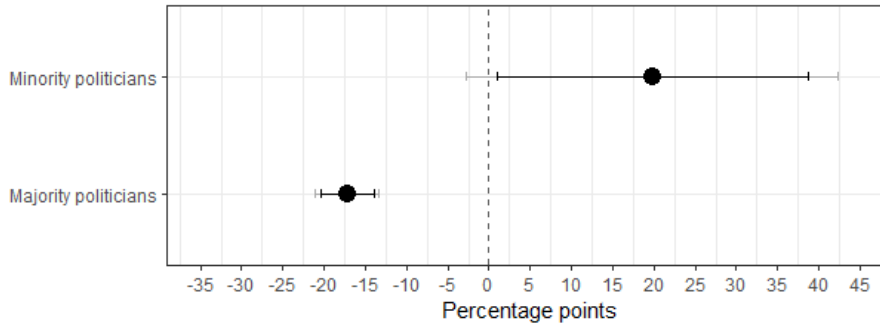
	<i>Dependent variable:</i>	
	Response (0/1)	
	(1)	(2)
Minority alias	-0.162*** (0.019)	-0.172*** (0.020)
Minority politician		-0.182** (0.085)
Minority alias * Minority politician		0.371*** (0.117)
Constant	0.709*** (0.014)	0.713*** (0.014)
Observations	2,395	2,395
R ²	0.028	0.032
Adjusted R ²	0.028	0.031

Note: The dependent variable in models 1-3 is receiving a reply (coded as 0/1). Coefficients are reported as percentages. White standard errors are reported in parentheses.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

¹⁰ For comparison, Butler (2014) finds that white legislators are 6.4 percentage points more responsive to white constituents, while black legislators was 3 percentage points more responsive to the black alias. Butler and Broockman (2011) find that minority constituents are approximately 10 percentage points less likely to receive a reply from majority politicians, while black politicians are 16.5 points more likely to reply.

Figure 2. Effect of the ethnic minority alias among ethnic majority and minority politicians



Note: The figure depicts the effect of the ethnic minority alias compared to the majority alias across incumbents' ethnic background. Bars indicate 90% and 95% confidence intervals.

Robustness of the findings

Before moving to the analyses of primary interest, we address – based on the baseline analyses – a range of general concerns pertaining to the successful implementation of the field experiment that may compromise our results more generally. A potential concern is that some incumbent politicians perceive the request as unrealistic. In particular, politicians from parties with an immigration-skeptical profile may be less likely to perceive the ethnic minority treatment as a genuine request. However, only two incumbents asked questions that could be interpreted as expressions of suspicion about the purpose of the treatment. Further, in Appendix G, we rerun the baseline analysis excluding the incumbents who are, arguably, most likely to perceive the treatment as dishonest – i.e., incumbents running for parties who are known for their immigration-skeptical profile – and find substantively similar results (ethnic minority alias = -14.7, $P < 0.001$).¹¹

Another concern is that incumbents from demographically small municipalities, as well as incumbents from municipalities with small ethnic minority populations, may be more inclined to perceive requests from immigrant-origin minority constituents to be unrealistic.¹² In other words,

¹¹ Specifically, we exclude Danish People's Party and the smaller party 'The New Right'.

¹² The smallest municipality in Denmark has less than 2,000 inhabitants, and there are 7 municipalities with less than 20,000 inhabitants.

legislators may perceive and respond to the emails differently due to their knowledge on the demographics of their constituents or simply because they know most of their constituents. To address this concern, we repeated the main analysis on subsets of the data across municipality size and find that the treatment effect is very consistent (See Appendix H) – the estimates range between 16.1 and 18 percentage points. Even when including only the seven smallest municipalities, where these concerns should be most pertinent, the estimated treatment effect of the ethnic minority alias is 16.8 (SE = 10.6) percentage points. We also explore if the bias against minority constituents is substantially stronger in municipalities with very few minority inhabitants, which is not the case (see details in Appendix I). A final concern involves interference between subjects. As noted earlier, the fact that the emails were sent over a given time interval reduces the risk that two incumbents received the mail at the same time, and from the replies, there is no sign that incumbents have discussed the requests. Taken together, our analyses of potential threats to the validity of our field experimental audit study indicate that it worked as intended, which therefore strengthens the credibility of our results.

Do electoral incentives reduce ethnocentric bias in responsiveness?

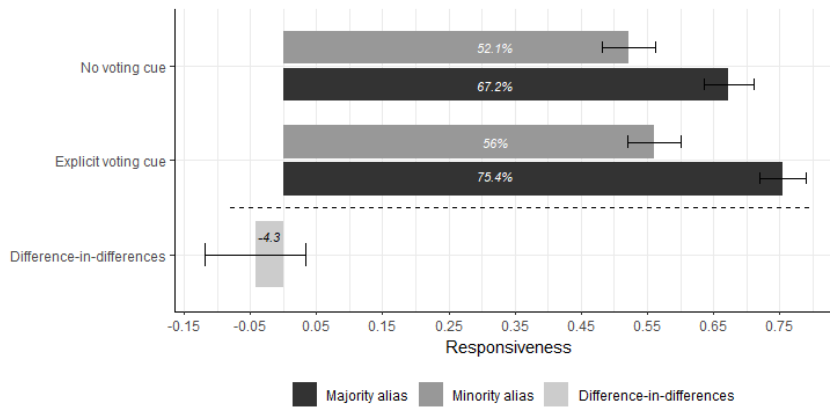
Having established that legislators exhibit strong ethnocentrism when responding to constituents, we now consider whether such bias can be curbed through the first proposed mechanism: electoral incentives for responding to any request in the interest of reelection. We test this proposition in a series of analyses probing majority legislators' responsiveness under increasingly strong electoral incentives.

First, to establish that variations in electoral competitiveness affect responsiveness more generally – and thus validate our manipulation of electoral incentives – we compared the response rate between baseline requests (without the stated intention to vote) and requests that included the explicit intention to cast a personal vote for the incumbent. The average treatment effect of the explicit personal voting intention increases the average response rate by 6 percentage points ($p < 0.01$), thus indicating that incumbents are responsive to this type of electoral incentive. We further validate this assertion in Appendix J, where we demonstrate that incumbents who seek re-election

are much more likely to respond (21.6 percentage points ($p < 0.001$) higher response rate) to the requests compared to incumbents in their final term.

Yet, our primary interest is whether this increase in electoral incentives equalizes differences in responsiveness between ethnic majority and minority constituents. To test this, we interact the stated vote intention with the ethnic minority alias. Figure 3 illustrates the primary result of this analysis (full results are reported in Table 2). It shows that the stated vote intention does not diminish majority legislators' ethnocentric responsiveness. In fact, in-group favoritism increases slightly under the explicit vote cue compared to requests without such cue, as indicated by a negative but statically insignificant interaction term (4.3 percentage points, $p = 0.28$).¹³

Figure 3. Effect of the personal voting cue across the ethnic alias-treatment



Note: Note: The Figure plots the average response rate to constituents with a majority or a minority alias conditional on whether a personal voting cue was included, as well as the difference-in-differences between these groups reported as percentage points ($N=2,326$).

While tapping electoral incentives in a very direct way, the relevance of the stated personal voting intention necessarily depends on the candidates' stake in the electoral outcome. For example, some incumbents hold safer seats than others and are therefore less likely to be animated by one

¹³ The difference-in-differences estimator compares the differences in response rates between in and out-groups when there is a clear voting relative to the difference when there is no voting intention.

extra potential vote. To strengthen the electoral incentive, in Table 2 we explore whether two observational variations in our data – politicians decision to stand for re-election and their winning margin in the 2017 election¹⁴ – decrease ethnocentric responsiveness when seen in conjunction with the stated vote intent. Although we cannot identify causal effects of these variations, they arguably serve to strengthen the electoral incentives for responsiveness.

We report the interaction between the ethnic alias and the personal voting cue in two subsets of incumbent politicians facing increasingly strong electoral incentives. For comparison, column 1 shows the interaction in the sample of majority politicians (as depicted in Figure 3). In column 2, we subset the data to include only majority incumbents seeking re-election ($N = 1993$), and finally, in column 3, we further restrict our sample so that it excludes all *clear winners*¹⁵ ($N = 492$). The overall result remains the same in both specifications; providing a personal voting cue does not reduce differential treatment. In fact, the interaction remains negative, implying that, if anything, strong electoral incentives prompt majority legislators to respond less to requests from minority aliases.

¹⁴ Recall that the experiment was conducted six weeks prior to the election; thus, at the time of the experiment, incumbents knew whether or not they were seeking re-election. Moreover, while a large share of the incumbents were certain to regain their seat in the local councils, defined here as clear winners, others risked losing their seat.

¹⁵ The clear winners win all of the 10,000 bootstrapped elections and therefore faced no serious risk of not getting elected.

Table 2. Interactions between Voting cue and Minority Alias

	<i>Dependent variable: Response</i>		
	All majority politicians	Reelection subset	Close election subset
	(1)	(2)	(3)
Minority alias	-0.151*** (0.028)	-0.153*** (0.030)	-0.193*** (0.061)
Vote cue	0.081*** (0.028)	0.089*** (0.029)	0.023 (0.062)
Minority alias * Vote cue	-0.043 (0.039)	-0.062 (0.042)	-0.021 (0.085)
Constant	0.673*** (0.020)	0.451*** (0.163)	0.440 (0.274)
Controls?	No	Yes	Yes
Observations	2,326	1,993	492

Note: The dependent variable is whether the incumbent politician responded to the request. All models estimated via OLS regression. *p<0.1; **p<0.05; ***p<0.01.

In summary, the results show that in-group favoritism in constituency service can occur even when there are clear strategic incentives for incumbents to respond to constituents. This is particularly striking given that at least two of the electoral incentives examined – receiving a request with an explicit vote intention in one’s favor and rerunning for election – strongly animates representatives to be more responsive to constituents on average. Yet, it does not constrain in-group favoritism.

Using party affiliation and policy preferences to identify responsive politicians

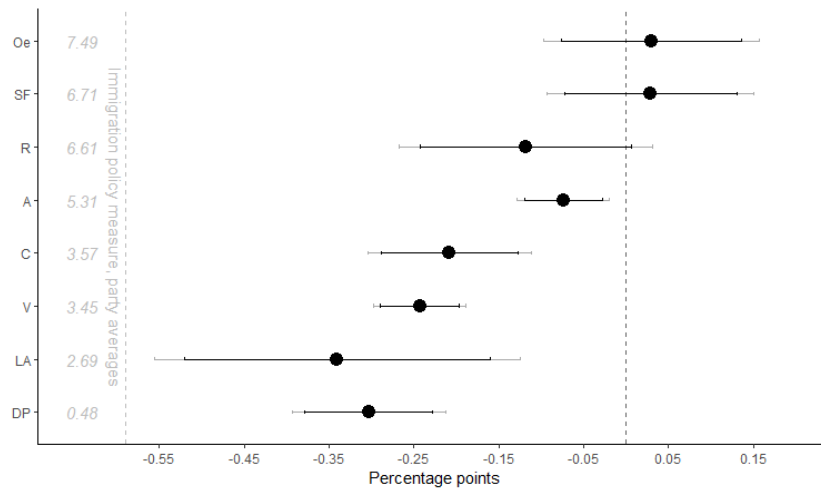
Unfortunate as it might be, the majority incumbents' lacking responsiveness to ethnic minority constituents even in the face of clear strategic incentives poses the question of what constituents can possibly do to identify legislators who are more responsive to them. This is particularly important in situations when voting for a co-ethnic political candidate – a solution following from the findings in the baseline model – is either not possible or undesirable.

The absence of electoral incentive-induced legislator responsiveness puts the burden on the constituents to identify parties and specific legislators that are responsive to them – what we label the candidate selection mechanism. More specifically, we examine whether immigrant-origin minority constituents – by using party affiliation and stated immigration preferences as heuristics – receive equal representation from parties and incumbents that explicitly articulate a motivation to promote the interest of their group. We do so in two steps: First, we examine whether ethnocentric responsiveness varies by political parties' positions on immigrants. Second, we test whether individual incumbent politicians' policy preferences moderate the effect of the ethnic minority alias once party differences are taken into account. To do this, we take advantage of data from the voting advice application as described in the Research Design section.

Figure 4 depicts the eight largest parties based on their candidates' average score on the immigration policy measure and plots the effect of the ethnic minority alias for each party along with 95 % confidence intervals. A lower score indicates that members of the parties on average articulate more restrictive immigration policies (with the exception of one party, the ranking closely follows how voters perceive parties' position immigration at the national level (Stubager et al. 2016)). The results suggest that using candidates' party affiliation can be quite effective in identifying legislators that are more responsive to minority constituents. Candidates from more immigration-friendly parties are much more responsive to minority requests than candidates from less restrictionist parties. In fact, incumbent politicians representing parties with the most liberal immigration policies are equally responsive to ethnic majorities and minorities. Among the two most left-wing parties,

the Unitarian list (Oe) and the Socialist People's Party (SF), the effect estimates are very close to zero. Among the two parties with the most restrictive immigration policy profiles, Danish People's Party (DF) and Liberal Alliance (LA), the estimated effects exceeds 30 percentage points. In the appendix, we present a number of alternative specifications that show that this result is robust to the inclusion of available covariates such as municipality size and share of immigrants (Appendix L).

Figure 4. Effect of the ethnic minority alias conditional on parties

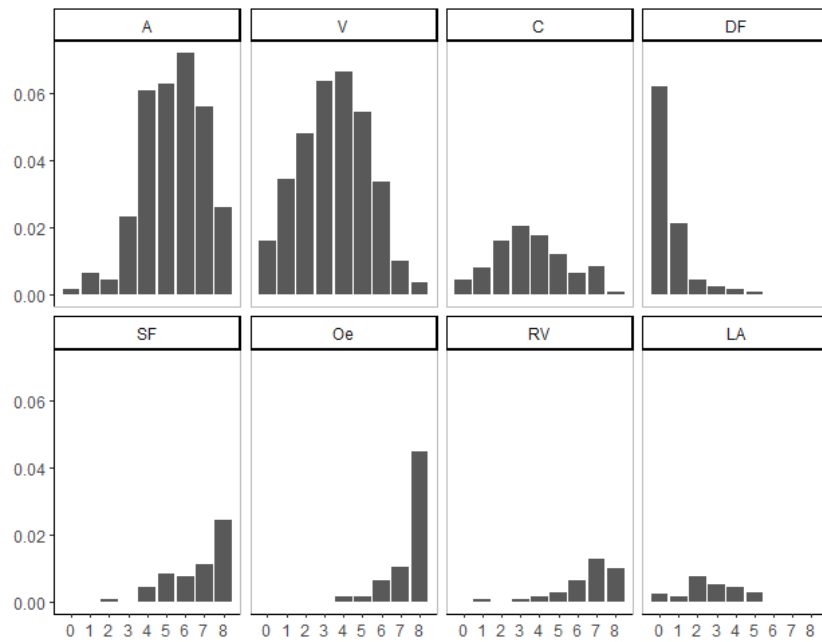


Note: The plot shows the effect of the ethnic minority alias among members of the eight largest parties. Parties are listed after their incumbent politicians' average score on the immigration policy measure (scores reported in italics). Scores of 0 and 8 indicate, respectively, the most and the least restrictive immigration policy profile.

For minority constituents, using party as a heuristic for obtaining equality in responsiveness is thus quite effective; contacting pro-immigration far-left parties secure parity in responsiveness. However, this strategy is not always possible or attractive to pursue. First, it may conflict with minority constituents' more general political preferences. Secondly, even within parties with a liberal immigration policy profile, there may be significant within-party variation in individual incumbents' preferences – especially within the larger catch-all parties. Figure 5 illustrates this point vividly;

while incumbent politicians' positions on the two (combined) minority policy questions are homogeneous within the smaller parties in local politics, there is substantial within-party variation among the three largest parties (the center-left Social-democratic Party (A), and the center-right Liberal Party (V) and the Conservative People's Party (C)). Within these parties there are both candidates who favour very restrictive policies, and candidates who favour very liberal immigration policies. This arguably reflects how local political candidates' often deviate from official party policies, but it also highlights that immigration policy is a divisive topic in Danish politics, especially in the large centrist parties. Therefore, incumbents' preferences towards policy questions on immigration are not necessarily conveyed by party affiliation.

Figure 5. Distribution of incumbents' position on the Immigration Policy Measure across parties



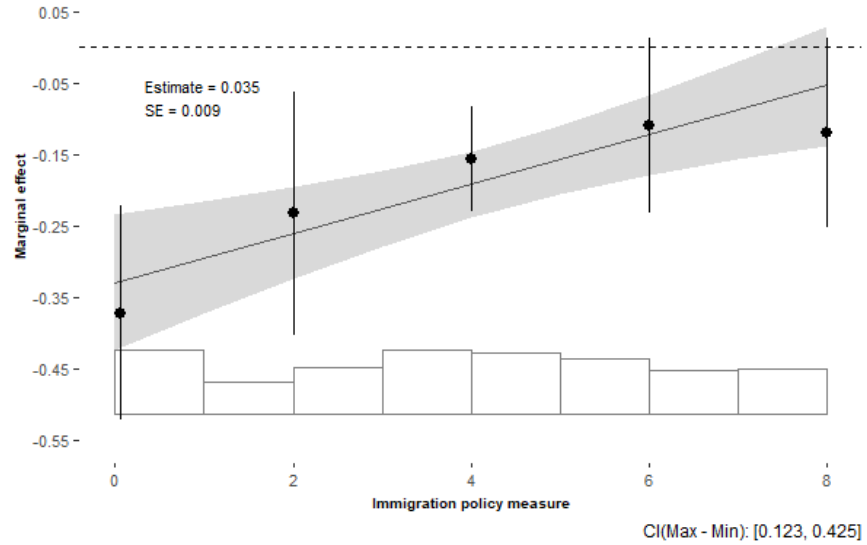
Note: The plots show the distribution of candidates' position on the immigration policy measure across political parties.

We therefore examine a second strategy, namely using individual incumbent politicians' policy preferences as a means to identify candidates that are more responsive to their requests even if they

are not running for parties with a clear pro-immigration profile. More precisely, we explore whether the relation between stated policy preferences on immigration issues and the effect of the ethnic minority alias persists once party affiliation is taken into account. Figure 6 plots the response-outcome regressed on an interaction between the minority alias treatment and the immigration policy measure with party fixed effects, that is, when analyzed within parties (the visual model include bins to assess the effect heterogeneity as suggested by Hainmueller, Mummolo, and Xu (2019)). As is evident from the plot, the conditional marginal effect changes at a near constant rate with the moderator. The interaction is positive and statistically significant (0.35; $p < 0.01$) indicating that, even when differences between parties are accounted for, personal preferences on issues of immigration are strongly associated with who gets a reply. This implies that using immigration preferences is a highly efficient heuristic for identifying responsive legislators for ethnic minority constituents. To put the effectiveness of the immigration preferences as heuristic in perspective, we examined interactions between the ethnic minority alias and each of the 13 remaining policy questions in the voting advice application. This analysis yielded one other significant interaction effect at the 0.01-level and two at the 0.05-level¹⁶ – all of these are for policy issues strongly correlated with answers to the immigration policy measure (Appendix K). This indicates that the preferences for immigration and integration policies specifically matters.

¹⁶ The question ‘The municipality should demand more from unemployed citizens’ yields a significant interaction effect at the 0.01-level. The two questions ‘The municipality spends too much on cultural activities’ and ‘Public institutions should serve ecological food’ yield significant interaction effects at the 0.05-level.

Figure 6. Marginal effect of the ethnic minority alias across immigration policy measure (+Party fixed effects)



Note: The plot shows the marginal effect of the ethnic cue across the immigration policy measure, including party fixed effects. In total, 1514 candidates answered both policy questions. A score of 0 on the measure indicates a restrictive immigration profile. Bins with 95% confidence intervals are included.

In summary, our results regarding the candidate selection mechanism suggest that incumbents' (parties') position on the immigration policy measure correlate strongly with the likelihood of answering the ethnic minority requester. On the one hand, this confirms conventional wisdom, that party affiliation is a strong indicator of individual candidates' policy preferences. On the other hand, however, it also reveals that the extent to which incumbent politicians' position on this question is conveyed through incumbents' party affiliation varies significantly. Thus, the analysis suggests that unless one sympathizes with parties on the far left, it is not enough to simply pay careful attention to party policies; constituents must be aware of individual candidates policy preferences to identify incumbents who are more likely to secure equal responsiveness. The broader implications of these results are discussed in the last section of the paper.

Conclusion and discussion

Studies have consistently demonstrated that political elites are more likely to provide various types of constituency service to members of their racial or ethnic in-group. This paper contributes to this literature by theorizing and testing two mechanisms that may further equal responsiveness. First, building on the assumption that politicians are extrinsically motivated in regard to (re)election, we explored whether electoral incentives animate instrumental legislators to be responsive to their constituents independent of the constituents' ethnic background. Second, we examined whether candidates' partisan affiliation and their stated policy preferences can be used as heuristics for identifying responsive legislators. To do so, we conducted a field experiment in which we randomized the traits of hypothetical constituents and contacted Danish local incumbents with a simple request for constituency service before the 2017 local elections.

The results indicate a large and precisely estimated negative effect on responsiveness of signing the requests to incumbent politicians with an immigrant-origin alias compared to requests signed with a majority name. Drawing on data on incumbents' electoral performance and a randomly assigned voting intention, we demonstrate that while incumbents who face strong electoral incentives devote more overall effort to constituency services, these incentives do not alter the bias in responsiveness. Differential treatment persists even when reelection-motivated politicians, who face close elections, receive a request stipulating a personal vote-intention. The results, however, provide strong empirical evidence for the second mechanism, namely that immigrant-origin constituents can identify legislators who are more responsive to them by paying attention to parties' and individual candidates' explicitly voiced policy positions vis-à-vis immigration.

These findings contribute to the empirical literature on in-group favoritism in politicians' responsiveness to constituents in several ways. First, scholars have argued that racial or ethnic biases in responsiveness are symptoms of a history of racial segregation and subjugation (McClendon 2016). By replicating the findings of unequal responsiveness from the United States and South Africa in the Danish setting, our analysis indicates that in-group favoritism is a more general phenomenon and not confined to specific racial or ethnic out-group constellations. Thus, the results suggest that

problems of inequality in responsiveness may be more widespread than previously thought.

Second, we extend the existing scholarship by demonstrating that immigrant-origin constituents can draw on parties and politicians' policy preferences to infer the likelihood of responsiveness. This finding opens up for a better understanding of the consequences of biases in politicians' constituency services. From one perspective, the finding is encouraging: if immigrant-origin constituents are selective when seeking constituency services from incumbents, they can obtain higher-quality responses. In fact, if they restrict communication to incumbents with the most pro-immigration policy positions, they receive political responsiveness to the same extent as their majority counterparts. From another perspective, the findings are discouraging. Specifically, the differential treatment remains significant across most parties, and thus a large share of local legislators – even those who favor pro-immigration policies – display discriminatory behavior. Moreover, voters often know little about individual candidates' policy positions (Carpini and Keeter 1996), and few citizens are motivated to learn the details of specific policy issues (Lupia, McCubbins, Arthur, et al. 1998). This leaves a large share of minority constituents with few tools to obtain the same level and quality of constituency communications. Or, put differently, they have to work harder to obtain the same constituency service as their majority counterparts.

The finding has larger institutional relevance too; as it turns out, constituents cannot simply rely on party cues as a heuristic to determine individual candidates' policy preferences. The significant within-in party variation in candidates' preferences on salient issues – and accordingly their correlated proximity of replying – emphasizes the importance of the possibility for voters to hold individual candidates accountable. This speaks in favor of electoral systems, such as the open-list PR system, in which the personal vote enables minority voters to enforce representation by electing specific candidates.

Our study leaves several questions to be explored in future studies. First, it is important to note that we cannot establish whether the link between explicitly stated support for a liberal immigration policy and ethnocentric responsiveness is due to strategic reasons or personal preferences. While this is not a problem for the proposed mechanism per se, future studies should illuminate this aspect further. Second, an avenue for future research is to test the viability of the candidate

selection-mechanism. As already highlighted, relying on detailed assessments of individual candidates' positions on specific policy issues may only be a viable strategy for a small share of the electorate who holds detailed knowledge about politicians' policy positions. Exploring the extent to which voters are able to follow this strategy is an important consideration in evaluating the effectiveness of the proposed mechanism. Finally, it is worth highlighting that further research is required to establish the robustness of our findings across different contexts.

The results have important implications for discussions about descriptive representation in Denmark, as well as in other Western democracies with substantial and growing immigrant-origin populations. A seminal definition of representation is 'acting in the interest of the represented in a manner responsive to them' (Pitkin 1967, p. 209). The gaps in incumbents' propensity to communicate with constituents, conditional on their personal characteristics, significantly affect equal access to the democratic process. Due to the composition of legislatures, this implies that racial or ethnic majorities are generally provided with better constituency service than are their racial or ethnic minority counterparts. This finding grants validity to the broader argument that descriptive representation influences how effectively minority groups receive equal substantive political representation.

References

- Alt, James, Ethan Bueno de Mesquita, and Shanna Rose (2011). “Disentangling accountability and competence in elections: Evidence from US term limits”. In: *The Journal of Politics* 73.1, pp. 171–186.
- Ariely, Dan, Anat Bracha, and Stephan Meier (2009). “Doing good or doing well? Image motivation and monetary incentives in behaving prosocially”. In: *American Economic Review* 99.1, pp. 544–55.
- Ashworth, Scott and Ethan Bueno de Mesquita (2006). “Delivering the goods: Legislative particularism in different electoral and institutional settings”. In: *The Journal of Politics* 68.1, pp. 168–179.
- Banducci, Susan A, Todd Donovan, and Jeffrey A Karp (2004). “Minority representation, empowerment, and participation”. In: *The Journal of Politics* 66.2, pp. 534–556.
- Bloemraad, Irene (2013). “Accessing the corridors of power: Puzzles and pathways to understanding minority representation”. In: *West European Politics* 36.3, pp. 652–670.
- Broockman (2013). “Black politicians are more intrinsically motivated to advance blacks’ interests: A field experiment manipulating political incentives”. In: *American Journal of Political Science* 57.3, pp. 521–536.
- Browning, Rufus P, Dale Rogers Marshall, and David H Tabb (1984). *Protest is not enough: The struggle of blacks and Hispanics for equality in urban politics*. University of California Press.
- Burden, Barry C. (2007). *Personal roots of representation*. Princeton University Press.
- Butler (2014). *Representing the advantaged: How politicians reinforce inequality*. Cambridge University Press.
- Butler and Broockman (2011). “Do politicians racially discriminate against constituents? A field experiment on state legislators”. In: *American Journal of Political Science* 55.3, pp. 463–477.
- Butler, Christopher F Karpowitz, and Pope (2012). “A field experiment on legislators’ home styles: service versus policy”. In: *The Journal of Politics* 74.2, pp. 474–486.

- Cain, John Ferejohn, and Morris Fiorina (1987). *The personal vote: Constituency service and electoral independence*. Harvard University Press.
- Canon, David T and Richard A Posner (1999). *Race, redistricting, and representation: The unintended consequences of black majority districts*. University of Chicago Press.
- Carnes, Nicholas (2012). “Does the numerical underrepresentation of the working class in Congress matter?” In: *Legislative Studies Quarterly* 37.1, pp. 5–34.
- Carpini, Michael X Delli and Scott Keeter (1996). *What Americans know about politics and why it matters*. Yale University Press.
- Coppock, Alexander (2016). “Randomizr: Easy to use tools for common forms of random assignment and sampling”. In: *R package version 0.5. 0*.
- Costa, Mia (2017a). “How responsive are political elites? A meta-analysis of experiments on public officials”. In: *Journal of Experimental Political Science* 4.3, pp. 241–254.
- (2017b). “Improving measures of responsiveness for Elite Audit Experiments”. In: *Working paper*. The working paper is found at <https://qssi.la.psu.edu/new-faces-papers-2017/costa-improving-measures-of-responsiveness-for-elite-audit-experiments>.
- Dahl, Robert A (1967). *Pluralist Democracy in the United States*.
- Dahlgaard, Jens Olav (2016). “You just made it: Individual incumbency advantage under proportional representation”. In: *Electoral Studies* 44, pp. 319–328.
- Dahlgaard, Jens Olav et al. (2009). “Kommunalpolitikeres rolle og råderum-2009”. In:
- Fenno, Richard F (1978). *Home style: House members in their districts*. HarperCollins,
- Fiorina, Morris (1989). *Congress: Keystone of the Washington establishment*. Yale University Press.
- Freeman, Patricia K and Lilliard E Richardson Jr (1996). “Explaining variation in casework among state legislators”. In: *Legislative Studies Quarterly*, pp. 41–56.
- Gell-Redman, Micah et al. (2018). “It’s all about race: How state legislators respond to immigrant constituents”. In: *Political Research Quarterly* 71.3, pp. 517–531.
- Grose, Christian (2011). *Congress in black and white: Race and representation in Washington and at home*. Cambridge University Press.

- Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu (2019). “How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice”. In: *Political Analysis* 27.2, pp. 163–192.
- Hemker, Johannes and Anselm Rink (2017). “Multiple dimensions of bureaucratic discrimination: Evidence from German welfare offices”. In: *American Journal of Political Science* 61.4, pp. 786–803.
- Hughes, D Alex et al. (2017). “Continuing Evidence of Discrimination Among Local Election Officials”. In: *Available at SSRN 3089893*.
- Iyer, Lakshmi et al. (2012). “The power of political voice: women’s political representation and crime in India”. In: *American Economic Journal: Applied Economics* 4.4, pp. 165–93.
- Janusz, Andrew and Nazita Lajevardi (2016). “The political marginalization of Latinos: Evidence from three field experiments”. In: *Available at SSRN 2799043*.
- Kotakorpi, Kaisa, Panu Poutvaara, and Marko Terviö (2017). “Returns to office in national and local politics: A bootstrap method and evidence from Finland”. In: *The Journal of Law, Economics, and Organization* 33.3, pp. 413–442.
- Lau, Richard R and David P Redlawsk (2001). “Advantages and disadvantages of cognitive heuristics in political decision making”. In: *American Journal of Political Science*, pp. 951–971.
- Lupia, Arthur, Mathew D McCubbins, Lupia Arthur, et al. (1998). *The democratic dilemma: Can citizens learn what they need to know?* Cambridge University Press.
- Mansbridge, Jane (1999). “Should blacks represent blacks and women represent women? A contingent” yes””. In: *The Journal of politics* 61.3, pp. 628–657.
- Mayhew, David R (1974). *Congress: The electoral connection*. Vol. 26. Yale University Press.
- McClendon, Gwyneth H (2016). “Race and responsiveness: An experiment with South African politicians”. In: *Journal of Experimental Political Science* 3.1, pp. 60–74.
- Mendez, Matthew (2014). *Who Represents the Interests of Undocumented Latinos? A Field Experiment of State Legislators*. Tech. rep. Working Paper, University of Southern California.
- Mendez, Matthew and Christian Grose (2018). “Doubling down: Inequality in responsiveness and the policy preferences of elected officials”. In: *Legislative Studies Quarterly* 43.3, pp. 457–491.

- Minta, Michael D (2009). “Legislative oversight and the substantive representation of Black and Latino interests in Congress”. In: *Legislative Studies Quarterly* 34.2, pp. 193–218.
- Nye, John VC, Ilia Rainer, and Thomas Stratmann (2014). “Do black mayors improve black relative to white employment outcomes? Evidence from large US cities”. In: *The Journal of Law, Economics, & Organization* 31.2, pp. 383–430.
- Pedersen, Lene Holm et al. (2013). *Lokalpolitikeres rolle og råderum*. KORA. Det Nationale Institut for Kommuner og Regioners Analyse og Forskning.
- Pitkin, Hanna F (1967). *The concept of representation*. Vol. 75. Univ of California Press.
- Popkin, Samuel L and Samuel L Popkin (1994). *The reasoning voter: Communication and persuasion in presidential campaigns*. University of Chicago Press.
- Siemiatycki, Myer (2009). “Reputation and Representation: Reaching for Political Inclusion”. In: *Electing a Diverse Canada: The Representation of Minorities and Women*, pp. 23–45.
- Sniderman, Paul M, Richard A Brody, and Phillip E Tetlock (1993). *Reasoning and choice: Explorations in political psychology*. Cambridge University Press.
- Stubager, Rune et al. (2016). “Danske vælgere 1971–2015”. In: *Det danske valgprojekt, Department of Political Science, Aarhus University*.
- Warshaw, Christopher (2019). “Local Elections and Representation in the United States”. In: *Annual Review of Political Science*.
- White, Ariel R, Noah L Nathan, and Julie K Faller (2015). “What do I need to vote? Bureaucratic discretion and discrimination by local election officials”. In: *American Political Science Review* 109.1, pp. 129–142.

Supplementary material for "Who is responsive? How electoral incentives and candidate selection shape ethnocentric responsiveness"

Appendix A. Aliases

Table A gives an overview of the 20 aliases included in the experiment to signal the putative ethnicity of the requester. Each last name was used twice and was paired with one male and one female first name. The names were chosen in order to resemble the most commonly used majority Danish and Non-Western names in Denmark. Average responses to each name is reported in parentheses. F-tests for the hypotheses that responses to candidates across names are equivalent (tested for majority names and minority names respectively) yields p-values of 0.65 and 0.48. As Email-addresses we used the treatment-name and a random number (e.g. JensenAnne78@gmail.com, AishaAbdi78@gmail.com)

Table A. Aliases

Majority alias				Minority alias			
<i>Male (response rate)</i>		<i>Female (response rate)</i>		<i>Male(response rate)</i>		<i>Female (response rate)</i>	
Peter Jensen	(0.68)	Anne Jensen	(0.70)	Mohammad Ibrahim	(0.54)	Hatice Ibrahim	(0.52)
Jens Nielsen	(0.69)	Kirsten Nielsen	(0.66)	Ahmad Omar	(0.50)	Fatma Omar	(0.58)
Michael Pedersen	(0.68)	Anna Pedersen	(0.69)	Mustafa Abdi	(0.58)	Aisha Abdi	(0.53)
Henrik Andersen	(0.71)	Hanne Andersen	(0.78)	Ali Mohamed	(0.48)	Amina Mohamed	(0.55)
Lars Hansen	(0.75)	Mette Hansen	(0.74)	Ibrahim Hussain	(0.63)	Fatima Hussain	(0.57)
H0: $P = 0.65$				$P = 0.48$			

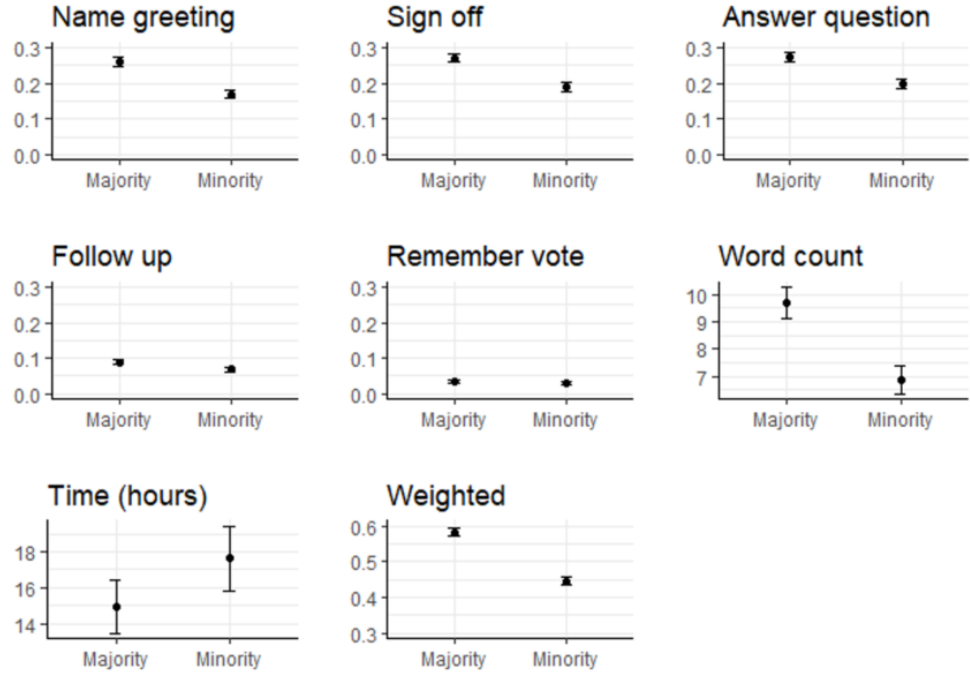
Note: Average response rates to each alias in parentheses

Appendix B. Content quality measures

In this appendix, we compare the quality of responses to majority and minority requesters using a number of qualitative measures as outcomes. Before reporting the results, we introduce the coding of the qualitative measures. Following the convention in the literature, we code all non-replies as 0 (Coppock, 2018; Holbein et al., 2018). In assessing the quality of the answers, we are inspired by the measure developed by Costa (2018). Specifically, we rely on seven content criteria, which we present individually and in a combined measure: *name greeting*, *invitation to follow up*, *answering the question*, *sign-off*, *length* (words beside greeting and sign off), *hours until response*. Costa’s measure includes the criteria no website link, which we exclude since it is unnecessary for answering the request. As a separate variable, we add an additional outcome by coding if politicians urged the requester to *remember to vote* at the election (or to vote immediately using the postal voting).

Figure C depicts the mean of quality of replies by treatment condition, with vertical lines representing 95% confidence intervals. Specifically, when the requester holds a minority name, the replies are significantly less friendly, less likely to provide an answer to the question, less likely to include an invitation to follow up, shorter, and less timely. We also calculate a weighted measure by discounting each measure with a given value (see Costa (2018) for details).

Figure B. Quality measures



Note: 95% confidence intervals reported around point estimates.

Appendix C. Details on simulation of electoral closeness

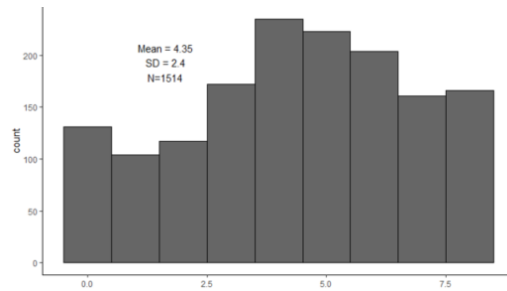
To measure individual candidates' electoral competition in the Danish PR-system, we used a bootstrapping-method inspired by Kotakorpi, Poutvaara, and Terviö (2017) and (Dahlgaard 2016). In short, we rely on the actual distribution of votes at the 2017 election to simulate a number of alternative elections for each incumbent to retrieve a distribution of simulations showing how often candidates re-win their seat. The procedure is essentially to (i) sample with replacement a vector of votes from the election in each municipality, and (ii) distribute votes to parties according to the votes in the sampled vector, before finally (iii) distribute seats to each candidate. In order to simulate uncertainty in who is elected, this process is repeated 10,000 times, which, for each candidate, creates a distribution of alternative elections in which individual candidates are either

elected or not. *Clear winners* are candidates who are elected every time (in all 10,000 simulations). Candidates who won (or lost) their seat with a small margin will be elected in fewer of the simulations. A majority of the incumbents are elected in all simulations, but a substantial share of candidates lose some of the simulated elections, thus indicating that they run in more competitive elections.

Appendix D. Distribution of immigration policy measure

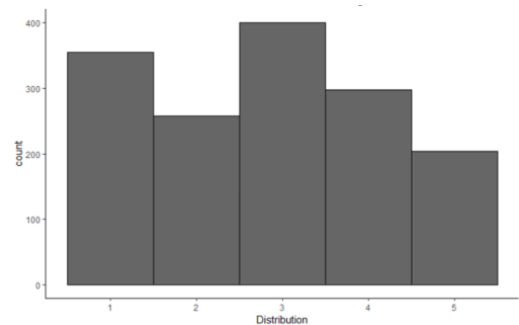
Figure D (i) depicts the distribution of incumbent candidates' answers when the two questions concerning immigration and integration are merged (as used in the main paper). Figure D (ii) and D (iii) give the distribution for each of the two questions.

Table D (i) Immigration policy measure



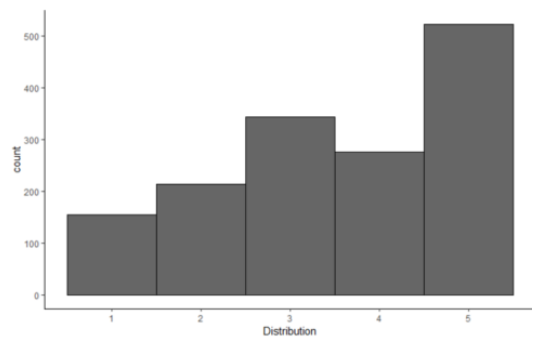
The plot shows the distribution of candidates' position on the immigration policy measure across political parties. Scores of 0 and 8 indicate, respectively, the most and the least restrictive immigration policy profile.

Table D (ii) The municipality should accept more refugees



The plot show the distribution of candidates' position on the question: '*The municipality should accept more refugees?*'

Table D (iii) Local municipal institutions try too hard to accommodate to religious minorities



The plot show the distribution of candidates' position on the question: '*Local municipal institutions try too hard to accommodate to religious minorities.*'

Appendix E. Effect of the gender cue

Table 1 indicates that the gender cue had no apparent effect on the likelihood of receiving a reply neither in general nor across incumbents' own gender.

Table E. Effect of the gender cue

	<i>Dependent variable:</i>
	Response
Female alias	-0.002 (0.024)
Female politician	0.018 (0.031)
Female alias * Female politician	0.029 (0.043)
Constant	0.619*** (0.017)
Observations	2,395
R ²	0.001
Adjusted R ²	-0.00005

Note: *p<0.1; **p<0.05; ***p<0.01.

Appendix F. Ethical considerations

Three concerns are especially worth paying attention to when conducting a field experiment like ours: (i) the burden imposed on experimental subjects; (ii) the time it takes to answer requests from the experiment takes time away from other citizens who may have been seeking help at the time; and (iii) the use of deception by relying on fictitious aliases. Several procedures were taken into account in order to keep the burden imposed on subjects at a minimum. First, to minimize the risk of misunderstanding and thus wasting incumbents' time, we carried out a small pilot study before fielding the experiment to ensure that the treatments were correctly understood and to test our experimental protocol. Secondly, we limited the time burden placed on each incumbent as much as possible. The emails sent in the context of our experiment were short and asked incumbents a fairly simple question that was straightforward to answer if an incumbent wished to do so. The median reply of about 30 words supports that this was also the case. Thirdly, to minimize any harm to the subjects, the replication data do not include any information making it possible to identify individuals. Similarly, data are not presented in any way that permits the identification of any individuals in the analyses. Accordingly, we believe that our experiment caused no significant harm to any of the incumbents or any constituents who may have been seeking help at the time where our experiment was conducted. Finally, the use of deception is particularly important to answer our research question. In order to study whether public legislators engage in discrimination, some degree of deception is unavoidable.

Appendix G. Are the treatments perceived as realistic?

Incumbents from parties running on highly restrictionist immigration policy platforms may have perceived the requests from minority constituents as unrealistic. In this appendix, we test if the main results hold up when excluding incumbents from these parties. As depicted in Table G, the overall differential treatment in responsiveness remains statistically significant in this subset ($N=2,094$), although the estimate is slightly smaller. Moreover, the interaction between the voting cue and the minority alias remains unaffected as compared to the analysis conducted in the full sample.

Table G. Restricted sample

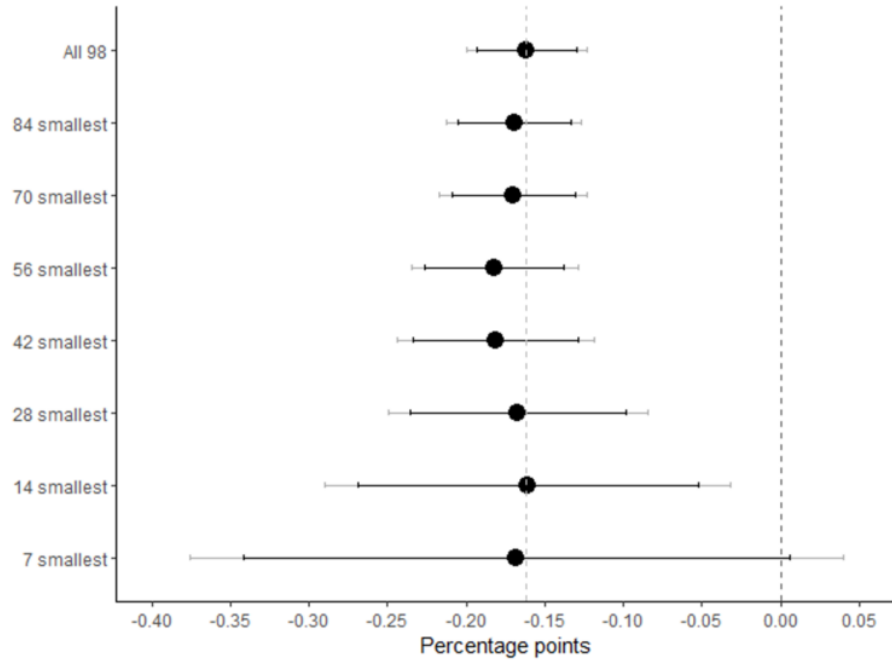
	<i>Dependent variable:</i>	
	Response	
	(1)	(2)
Minority alias	-0.147*** (0.021)	-0.126*** (0.029)
Voting cue		0.084*** (0.029)
Minority alias * Voting cue		-0.042 (0.042)
Constant	0.703*** (0.015)	0.662*** (0.021)
Observations	2,094	2,094
R ²	0.023	0.028
Adjusted R ²	0.023	0.026

Note: *p<0.1; **p<0.05; ***p<0.01.

Appendix H. Average treatment effects across municipalities listed by size

Incumbents from demographically small municipalities, as well as incumbents from municipalities with small ethnic minority populations, may be more inclined to perceive requests from immigrant-origin minority constituents to be unrealistic. Figure H depicts the ATE of the minority alias across subsets of the data listed after the number of inhabitants in the municipalities. The plot indicates that even when including only the seven smallest municipalities, where these concerns should be most pertinent, the estimated treatment effect of the ethnic cue is 16.8 (SE = 10.6).

Figure H. Average treatment effects across municipality size

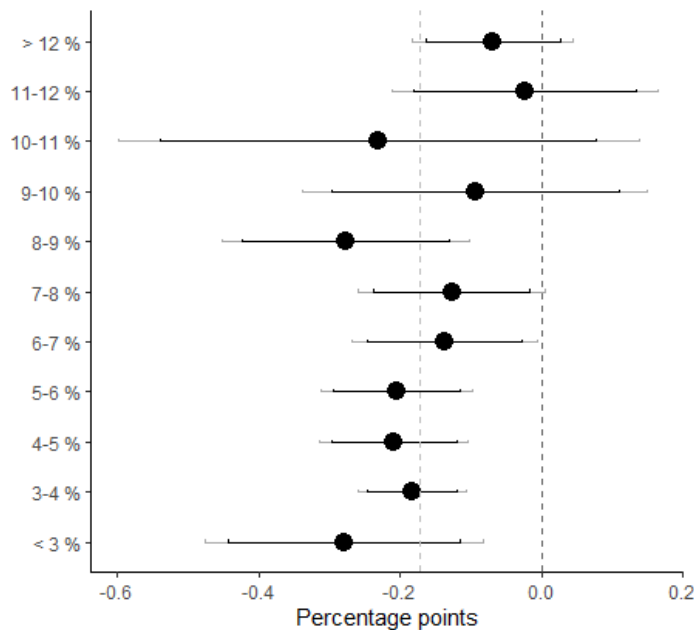


Note: Each estimate represents the ATE among incumbents in municipalities with a given size. The grey line indicates the average treatment effect. 90% and 95% confidence intervals reported around point estimates.

Appendix I. Average treatment effects across the share of immigrant-minority inhabitants in municipalities

Figure J depicts the ATE of the minority alias across subsets of data arranged after the share of immigrant citizens living in the incumbents' municipalities. While the ATE is larger than average in municipalities with few ethnic minority citizens (although estimated with substantial uncertainty), there is no clear empirical pattern.

Figure I. Average treatment effects across share of immigrants



Note: Each estimate represents the ATE among incumbents in municipalities with a given share of immigrant citizens. The grey line indicates the average treatment effect. 90% and 95% confidence intervals reported around point estimates.

Appendix J. Electoral incentives shape responsiveness

This appendix examines how contextual variations in electoral incentives affect incumbency responsiveness more generally, and whether they themselves moderate the effect of the ethnic minority alias. The electoral incentive from running for re-election does not appear to mitigate the effect of the ethnic cue. As seen in column 2, the effect of the minority cue appears, in fact, to be larger among incumbents that rerun, although statistically insignificant.¹⁷ In the same vein, column 3 introduces a dichotomous variable indicating if incumbents were clear winners (=1) or if they faced a risk of losing their seat. There is no overall effect of incumbents' level of electoral uncertainty nor does it moderate the effect of the minority alias.¹⁸

Figure J. OLS regression results across variations in electoral incentives

	<i>Dependent variable: Response</i>		
	(1)	(2)	(3)
Minority alias	-0.172*** (0.019)	-0.149*** (0.026)	-0.211*** (0.033)
Rerun	0.157*** (0.018)	0.181*** (0.026)	
Minority alias * Rerun		-0.049 (0.036)	
Clear winner			0.022 (0.031)
Minority alias * Clear winner			0.042 (0.043)
Constant	0.415*** (0.147)	0.412*** (0.147)	0.491*** (0.164)
Controls?	Yes	Yes	Yes
Observations	2,326	2,326	1,888

Note: *p<0.1; **p<0.05; ***p<0.01.

¹⁷ There are several potential explanations for this result. A likely explanation is that the final term-incumbents that do provide an answer are relatively more intrinsically motivated to provide constituency service, which crowds out part of the effect of the ethnic minority alias. It could also be that re-running incumbents are busier and thus put more weight on strategic considerations about the likelihood of who votes.

¹⁸ As suggested by Butler, Pope and Karpowitz (2012), causality may also go in the other direction: Legislators who fail to prioritize service do not do as well during elections.

Appendix K. Voting advice application

The voting advice application includes the following questions:

- Q1: Private companies should be in charge of a larger part of the elderly care
- Q2: Municipalities should involve volunteers to help employees at nursing homes
- Q3: The municipality should lower taxes
- Q4: The municipality can save money without a decrease in welfare services
- Q5: The municipality should ensure socially diverse schools
- Q6: Municipalities should set a limit to the number of kids per pedagogue in kindergartens
- Q7: The municipality should spend more money on bicycle lanes
- Q8: The municipality should accept more refugees
- Q9: Local municipal institutions try too hard to accommodate to religious minorities
- Q10: The municipality spends too much on libraries and cultural institutions
- Q11: Elderly who can afford it should be able to choose additional services at nursing homes
- Q12: The municipalities should spend more on sorting and reusing trash
- Q13: The municipality should demand more from unemployed citizens
- Q14: Children spend too much time in schools
- Q15: Public institutions should serve ecological food

Table K (*i*) and (*ii*) shows the results from regressing the response outcome on the ethnic treatment, each of the individual questions included in the voting advice application and the interaction term

Table K (i)

	Response							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Minority alias	-0.326*** (0.049)	-0.179*** (0.046)	-0.165*** (0.051)	-0.106** (0.047)	-0.094* (0.056)	-0.268*** (0.081)	-0.250*** (0.078)	-0.312*** (0.101)
Immigration policy measure	-0.006 (0.007)							
Minority alias: Immigration policy measure	0.034*** (0.010)							
Q1		-0.023* (0.012)						
Minority alias:Q1		0.001 (0.018)						
Q2			-0.014 (0.012)					
Minority alias:Q2			-0.005 (0.018)					
Q3				0.003 (0.011)				
Minority alias:Q3				-0.027* (0.016)				
Q4					-0.005 (0.011)			
Minority alias:Q4					-0.026 (0.016)			
Q5						-0.009 (0.014)		
Minority alias:Q5						0.023 (0.020)		
Q6							0.011 (0.013)	
Minority alias:Q6							0.018 (0.019)	
Q7								0.003 (0.016)
Minority alias:Q7								0.032 (0.024)
Constant	0.799*** (0.034)	0.825*** (0.032)	0.811*** (0.036)	0.766*** (0.033)	0.788*** (0.039)	0.807*** (0.056)	0.732*** (0.054)	0.761*** (0.069)
N	1,513	1,513	1,513	1,513	1,513	1,513	1,513	1,513
R ²	0.047	0.041	0.039	0.040	0.041	0.037	0.040	0.039

Note: *p<0.1; **p<0.05; ***p<0.01.

Table K(ii)

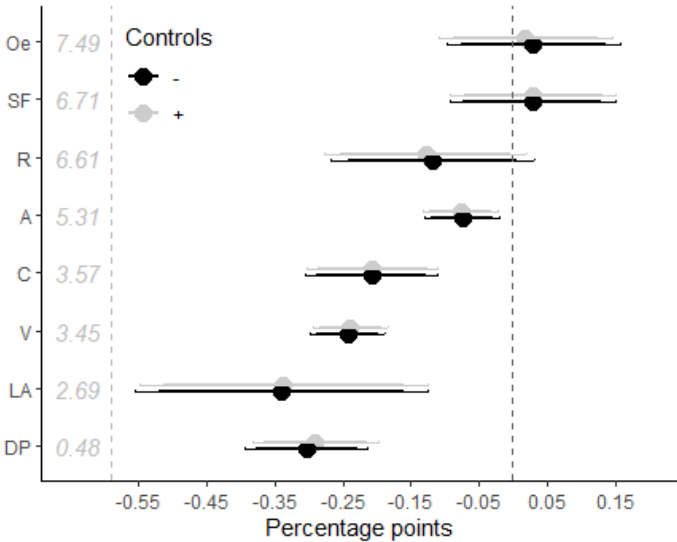
	Response						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Minority alias	-0.326*** (0.049)	-0.074 (0.048)	-0.070 (0.062)	-0.322*** (0.100)	-0.010 (0.059)	-0.078 (0.069)	-0.270*** (0.050)
Q10		0.008 (0.015)					
Minority alias:Q10		-0.052** (0.021)					
Q11			-0.009 (0.012)				
Minority alias:Q11			-0.031* (0.016)				
Q12				-0.0003 (0.016)			
Minority alias:Q12				0.034 (0.023)			
Q13					-0.011 (0.012)		
Minority alias:Q13					-0.051*** (0.017)		
Q14						0.017 (0.013)	
Minority alias:Q14						-0.028 (0.018)	
Q15							0.006 (0.011)
Minority alias:Q15							0.033** (0.016)
Constant	0.799*** (0.034)	0.758*** (0.033)	0.803*** (0.044)	0.775*** (0.068)	0.809*** (0.041)	0.714*** (0.049)	0.758*** (0.036)
N	1,513	1,513	1,513	1,513	1,510	1,513	1,513
R ²	0.047	0.042	0.044	0.039	0.054	0.038	0.044

Note: *p<0.1; **p<0.05; ***p<0.01.

Appendix L. Effects across parties

In the paper, we demonstrate that incumbent politicians representing parties with the most liberal immigration policies are equally responsive to ethnic majorities and minorities. Table L presents the results across parties with and without controls (municipality size and share of immigrant-origin citizens on the municipality-level).

Figure L. Average treatment effects across parties



Note: The Figure depicts ATE across parties with and without controls. 90% and 95% confidence intervals reported around point estimates.

Candidate choice in a high-information setting: Do ascriptive characteristics shape candidates' electoral prospects?

Malte Dahl and Jacob Nystrup*

Working paper presented at DPSA 2018

Abstract

Research shows that voters rely on political candidates' ascriptive characteristics as heuristics to infer personality traits or particular policy positions. This suggests that characteristics such as gender or ethnicity may shape candidates' electoral prospects. However, we know little about whether these characteristics matter when voters have abundant information about the candidates, which is often the case in elections. In this short research note, we examine voter preferences towards two politically under-represented and salient social groups: women and immigrant-origin minorities. In a candidate choice conjoint experiment, we asked a representative sample of voters ($N=1,551$) to choose between two hypothetical local political candidates. Voters were presented with detailed candidate information including party membership, policy position on salient topics and political experience. We find a significant bias against candidates with immigrant-origin names and show that the effect is remarkably consistent across voter characteristics. We find no pro-male bias in general or in combination with other candidate traits.

*Malte Dahl is a PhD candidate in the Department of Political Science, University of Copenhagen, 1353 Copenhagen (mrd@ifs.ku.dk). Jacob Nystrup is a PhD candidate at Nuffield College, University of Oxford, (jacob.nystrup@nuffield.ox.ac.uk)

Introduction

The under-representation of immigrant-origin minorities and women in the political sphere is a pervasive feature of most democracies (Dancygier et al. 2015; Teele, Kalla, and Rosenbluth 2018). Denmark, the focus of this paper, fits this trend. In recent decades, the share of women in city councils has remained fairly constant, at around one-third of seats. Immigrant-origin minorities hold only 3% of the seats in the local councils though they comprise 7% of the electorate. To understand this under-representation, a central question is the extent to which these gaps are related to an unwillingness on the part of the electorate to vote for certain types of candidates – so-called electoral discrimination (Portmann and Stojanović 2019).

With regard to the under-representation of women, there is some debate about the extent to which gender stereotypes shape voter preferences. While gender stereotypes have been found to hamper female candidates' electoral prospects in subtle ways (Teale, Kalla, and Rosenbluth 2018; McDermott 1998), an increasing number of studies seem to indicate that the under-representation of women does not result from a direct preference for male candidates (Schwarz, Hunt, and Coppock 2018; Kage, Rosenbluth, and Tanaka 2017; Lawless and Pearson 2008). In comparison to the vast literature on gender, the body of evidence on voter preferences for immigrant-origin candidates is scarce. Recent studies of electoral outcomes in the U.K. (Thrasher et al. 2017; Fisher et al. 2015) and Switzerland (Portmann and Stojanović 2019) suggest that immigrant-origin candidates face barriers in the election process. This aligns with work that considers the U.S. context (Philpot and Walton Jr 2007; McConnaughy et al. 2010; Besco 2015; Broockman and Soltas 2017).

However, methodological challenges present difficulties in exploring the factors that influence voters' political candidate preferences. Because candidates run in different districts, on different platforms, for different positions, with different experience and resources, studying the make-up of legislatures provide little information about the characteristics that affect voter demands (Broockman and Soltas 2017). Moreover, the fact that relatively few immigrant-origin candidates run for elections amplifies the difficulties of examining the electoral support for this group based on observational data (Dancygier et al. 2015). Moreover, survey experiments that explore the effects of

candidates' ascriptive traits – immutable characteristics such as gender or ethnicity – have typically done so in low-information settings where most information beside these traits is held constant (Campbell and Cowley 2014; Banducci et al. 2008; Bauer 2015; Mo 2015; McConaughy et al. 2010; Philpot and Walton Jr 2007). While there is ample evidence to indicate that voters use candidates' ascriptive traits as information shortcuts to make inferences about candidates' policy preferences (Cutler 2002; Sen 2017; Atkeson and Hamel 2018) or personality traits (Carnes and Lupu 2016; Kirkland and Coppock 2017; Pedersen, Dahlgaard, and Citi 2019), it is less certain if ascriptive traits matter when voters have access to detailed candidate profiles. In other words, are female and immigrant-origin candidates disadvantaged due to voter biases when voters have little reason to use these descriptive characteristics as heuristics?

To address this question, we invited a nationally-representative sample of voters to participate in multiple 'elections' in form of a candidate choice conjoint experiment. In this experimental setting, we mimicked real-world elections by asking voters to evaluate multiple pairs of hypothetical local political candidates described by several randomly assigned attributes that have been identified as being relevant to a personal vote. We provided ample information about the candidates such as age, job, political experience and party affiliation, and their positions on two salient questions regarding economic and social policies. This approach builds on a small but increasing number of conjoint experiments used to evaluate how voters react to candidate attributes (e.g. Teele, Kalla, and Rosenbluth (2018), Carnes and Lupu (2016), Sen (2017), and Ono and Yamada (2016)). To our knowledge, this is the first candidate choice conjoint experiment designed with the express purpose of exploring voter preferences regarding immigrant-origin minorities and female candidates in a European context.

Our conjoint experiment enables us to uncover whether voters – both on average and across subgroups – exhibit preferences that benefit certain types of candidates directly or in more subtle ways. A key value of the experiment is the possibility to explore voters' multidimensional preferences simultaneously, which holds some important advantages for the inferences that we are able to draw. Our design strengthens external validity by providing voters with information that they often have access to, and base their decision-making on, in elections. Additionally, by randomly assigning

all attributes we can compare the relative importance of candidates' ascriptive traits with the importance of other characteristics. Moreover, we are able to explore whether the characteristics that voters reward are conditional on these traits. Finally, the high-information setting makes our survey experiment a test of whether ascriptive traits matter when there are limited reasons to use them as a heuristic.

The results indicate that political candidates with an immigrant-origin alias are significantly disadvantaged by voters. Thus, having an immigrant-origin name appear to be a *deal-breaker heuristic* that voters use to eliminate candidates (Bernhard and Freeder 2018). Voters on average favour the ethnic majority candidate by five percentage points over candidates with an immigrant-origin alias. Further, this treatment effect is consistent across various subsets of the sample: ethnic majority candidates are preferred across voter characteristics including gender, political interest and, to some extent, political leaning. Moreover, voters appear to evaluate candidates' other characteristics differently when the candidates have an immigrant-origin background as compared to how ethnic majority candidates are evaluated. Generally, immigrant-origin minority candidates are rewarded more for having a job that requires a lengthy education, but punished for preferring less socially diverse schools. The experiment also shows that female candidates do not operate at a disadvantage – if anything voters on average favour female candidates by two percentage points. Moreover, when interacting gender with other traits, we find no signs of double standards: otherwise identical candidates are evaluated very similarly across gender. Finally, we find no evidence for the notion that fewer female candidates hold beneficial traits compared to men in the actual pool of candidates.

The context: descriptive representation in Danish municipal politics

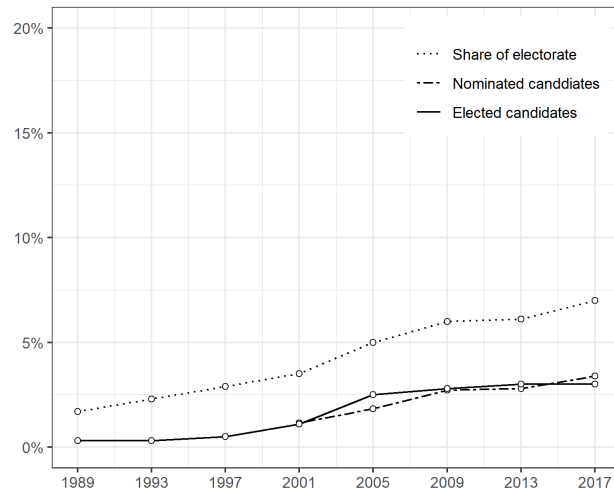
Denmark is characterized by a large degree of decentralization. The municipalities are responsible for around 50% of the total government expenditure, which is the highest in Europe (Eurostat 2017). This budget is spend on a long range of public services such as schools, local infrastructure and

social benefits. Furthermore, the local city councils enjoy a large degree of financial and political autonomy. Local politics is therefore hugely important in Denmark and, as a result, local elections are heavily contested and enjoys much attention from parties, the media and voters. Turnout has consistently been high; in the local election in 2017, the national turnout was 70.6%.

There are 98 municipalities in Denmark, which ranges from around 2.000 inhabitants to more than 600.000. Each municipality has an elected city council. The smallest consists of 9 persons (Læsø), while the largest consists of 55 persons (Copenhagen). The members of the city councils are up for re-election every four years. The election system is proportional, and everyone can—as long as they have a limited number of signatures—run for election. The barriers to running are low, and in the latest election around 1 in 400 voters were running for a seat in the city council. In 2017, 9,558 candidates were competing for 2,432 seats in the city councils. As seen in Figure 1, 3% of the candidates were immigrant-origin minorities, which is slightly less than the share of nominated candidates and less than half of the electorate of 7%. Figure 2 shows that 31.8% of the candidates were women, while 33.0% of the elected members of the city council were women. This indicates that among the pool of candidates, women were slightly more likely to be elected given they were running as a candidate.¹

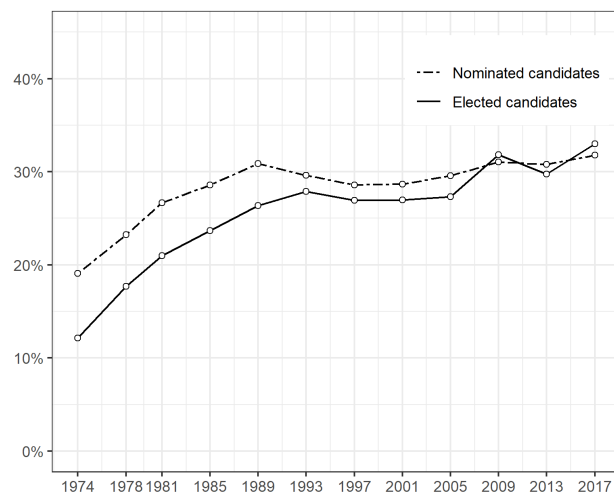
¹ As seen in figure 2, it has not always been the case that women were more likely to be elected. In 1974, a male candidate was almost 60% more likely to be elected than a female candidate. The share of female candidates running for office increased drastically in the 1970'ies and 80'ies, but stagnated in 1989 and has essentially remained constant ever since (Kjær 2013).

Figure 1. Share of immigrant-origin minorities in Danish Municipal Politics



Source: Togeby (2008) and Statistisk årbog

Figure 2. Share of women in Danish Municipal Politics



Source: Statistisk Aarbog

Candidate characteristics and voter preferences

There are several accounts for the under-representation of immigrant-origin minorities and women in legislatures. Prominent explanations include barriers within political parties (Norris and Lovenduski 1995), a lack of resources (Dancygier et al. 2015; Brouard and Tiberj 2010), the incentives established by electoral institutions (Fox and Lawless 2010), that different political issues spur interest among men and women (Carroll and Sanbonmatsu 2013; Bækgaard and Kjær 2012) and that ambitions and self-perceived qualifications differ across gender (Shames 2017; Fox and Lawless 2010). This paper specifically focuses on whether voters exhibit preferences based on candidates' ascriptive traits when selecting between otherwise identical candidates – so-called electoral discrimination.² The following sections briefly review the literature on how voter preferences are shaped by political candidates' ethnic affiliation and gender.

Immigrant-origin minority candidates and voter demands

While ethnic minorities are under-represented in the political bodies of most democracies, the explanation behind this under-representation has received much less empirical attention than the question of gender (Thrasher et al. 2017; Street 2014). As underlined by Broockman and Soltas (2017), ample evidence indicates that individuals engage in racial and ethnic discrimination in product and labor markets, but little work has assessed whether they do so when voting in elections.

Some studies indicate that non-native candidates seem to perform worse than majority candidates on election day. A study of the 2010 general election in the U.K. identified that incumbents could expect to gain approximately two percentage points in voter share of if they ran against an ethnic minority candidate, and even more if two minorities appeared on the ballot (Stegmaier, Lewis-Beck, and Smets 2012). In the same vein, Fisher et al. (2015) found that British voters are

² Albeit the term *electoral discrimination* is commonly used (Portmann and Stojanović 2019; Thrasher et al. 2017), it is not entirely clear if electoral preferences over ethnicity equals discrimination per se. In the context of local council elections, Thrasher et al. (2017) argues that 'it may be that some voters prefer to choose candidates that best fit the dominant demographic characteristics of the area – a case of voting for 'people like me'. However in the broadest sense, it is entirely valid to address concerns in the wider society about the reaction of some voters to candidates based on their ethnic identity, because this affects the nature of representative democracy.

less willing to vote for ethnic minority candidates, especially Muslims. This aligns with survey experiments that indicate that voters often have a preference for voting for a co-ethnic (Philpot and Walton Jr 2007; McConaughy et al. 2010; Besco 2015).

There are multiple mechanisms underlying why voters might hold negative biases against ethnic out-group candidates. One possibility is that candidates' ascriptive traits serve as heuristics – the use of a cognitive shortcut to simplify opinion formation – indicating the candidate's party affiliation, values or ideology (Kirkland and Coppock 2017; McDermott 1998; Sanbonmatsu 2002). Voters can often reliably assess candidates' gender and ethnic background based on the name printed on the ballot, information that also serves as a shortcut for partisan affiliation, political experience or issue positions (Huddy and Terkildsen 1993; Philpot and Walton Jr 2007). Thus, voters 'extrapolate stereotypical information' and use it to guide voting decisions (McDermott 1998). For example, in the U.S. context, women and African-American political candidates are seen as more socially liberal than white men (Huddy and Terkildsen 1993; McDermott 1998).

Another potential mechanism is that voters may evaluate candidates using a likability heuristic reflecting their affect toward salient groups. Voters may hold a 'distaste' for certain groups. This personal preference implies a psychic cost related to casting a vote for an out-group candidate (Broockman and Soltas 2017). In a related vein, voters might favour in-group members due to a feeling of group loyalty (Fisher et al. 2015). Finally, voters may believe that candidates from different ethnic groups are not equally qualified for the job. Fisher et al. (2015) draws on an analogy regarding 'statistical discrimination' in the labour market, whereby employers discriminate on the basis of assumed productivity differences between groups. Thus, because voters have limited information about candidates when voting, they make inferences about the candidates' qualifications based on their ethnic group.

Female candidates and voter demands

Abundant research has demonstrated that voters stereotype political candidates based on gender. A widely accepted view is that voters perceive men as more competent and assertive, and women as more compassionate, warm and emotional (Dolan 2014; McDermott 1998; Huddy and Terkildsen

1993). By the same logic, women are assumed to be better suited to specific policy areas such as general welfare, social services and the environment, while men are believed to have more skill dealing with economic development, trade, taxes and agriculture (Dolan 2010; Brown, Heighberger, and Shocket 1993; Huddy and Terkildsen 1993).

Despite the evidence on gender stereotypes, an increasing body of research finds little support for a direct voter bias against female candidates. In the U.S. political context, many scholars argue that ideology or partisanship are the main determinants of voting behaviour, details more important to candidate choice than gender (Dolan 2014; Lawless 2015). Moreover, a second line of research suggests that women are now perceived to be as competent politicians as men (Carnes and Lupu 2016; McElroy and Marsh 2010). When considering survey experimental research on voter preferences, a majority of studies in fact show a small net preference for women (see Schwarz, Hunt, and Coppock (2018) for a review).

Even though the evidence does not indicate a direct disadvantage for female candidates, there are ways that gender could still play a role in shaping electoral outcomes. For example, voters may apply so-called double standards by evaluating traits differently for men and women (Teele, Kalla, and Rosenbluth 2018). Moreover, certain traits that voters pay attention to may characterize male and female candidates with different frequency. We can imagine that voters prefer certain demographic traits – e.g. candidates who have more experience, are older and well-educated – that may fit a larger share of male candidates.³ In the present study, we therefore consider four potential ways that female candidates may be disadvantaged due to voter preferences. First, we examine if voters have a direct preference for male over female candidates. Secondly, we consider this across a number of voter subgroups. Thirdly, we explore if the effects of other traits are conditioned on the candidates' gender – in other words, we examine if voters apply double standards. Fourthly, we examine how demographic traits that affect voters' choice of candidates are distributed across gender in the actual pool of candidates.

³ This is related to the notion of 'double-ties' described by Teele, Kalla, and Rosenbluth (2018) as '*when desirable traits require more investment, or are associated with different burdens' for certain groups*'.

Design and data

The survey experiment was conducted in a commercial web panel (AnalyseDanmark), where members were invited by email to participate. A total of 2,123 respondents started the survey, and 1,551 completed it in its entirety, which corresponds to a completion rate of 73%. The sample is approximately representative of the adult Danish population in terms of gender and age, while the educational level in the sample was somewhat lower compared to the general population.⁴ Respondents' self-placement on a political left-right scale indicated an acceptable level of ideological variation (Mean = 4.8, SD=2.4, scale 0-10). For a comparison of sample and population characteristics, see Appendix A. Male and female respondents were well-matched in age, geography and likelihood to vote in recent elections.

Candidate choice conjoint experiment

At the beginning of the survey, respondents were asked questions on demographics and vote choice, and invited to place themselves on a political left-right scale (0-10). Respondents were then exposed to the candidate choice conjoint experiment.

The experiment is a paired-conjoint design in which respondents are asked to choose between two hypothetical political candidates described with seven attributes that could each hold two or more levels. Compared to a single-attribute survey experiment, our conjoint experiment resembles a more realistic scenario in which voters are able to consider candidates that differ on a variety of dimensions, which accordingly increases the external validity (Hainmueller, Hopkins, and Yamamoto 2014). Each respondent was presented with five pairs of side-by-side profiles of randomly generated candidates (see Appendix B for an illustration of these profiles) and a forced-choice between two candidates. By asking each respondent to evaluate several pairs of candidates, we increase the effective sample size of the survey (7,755 choices in total). The experiment was implemented using Qualtrics software.

The following candidate attributes were randomly assigned: name, age, profession, political

⁴ Age in years in the sample $M = 53$, $SD = 16.6$ ctr. population $M = 49.1$, $SD = 18$. The share of female: sample = 49.7% ctr. population = 50.8%

experience, position on two policy questions and party affiliation. The names that were used represent four different categories across gender (male/female) and ethnic affiliation (majority/minority). Contrary to most conjoint experiments, we used names as proxies for gender and ethnic affiliation, since providing information about ethnicity is uncommon in a Danish context. Arguably, this maximizes realism and circumvents the anticipation of respondents for the two attributes ‘gender’ and ‘ethnicity’ as being specifically important, which could result in response bias. By the same logic, we restricted the probability of any given candidate having an ethnic minority name to 20%. In order to cue immigrant-origin background, we relied on a range of highly popular names that are easy distinguishable to provide precise signals for ethnicity and gender.⁵

In order to avoid the choice of candidates based on inferences about the connection between gender or ethnicity and party-affiliation we included partisanship as an attribute (as is also the case on actual ballots). However, to avoid that partisanship mutes effects of other traits (Carnes and Lupu 2016), we held the candidates’ party label constant in each pair, i.e. both candidates were affiliated with either the Liberal Party or the Social Democrats.⁶ We also provided candidates’ positions on two policy questions to mitigate any inference voters may make regarding candidates’ policy preferences based on ethnicity or gender. We relied on an economic question ‘*Should local taxes be reduced?*’ and a question on social policy ‘*Should municipalities secure diversity in schools?*’. The questions were adopted from the ‘Candidate Test’, a popular online voting advice application of 15 policy questions that local politicians answer before the election, making it possible for voters to match their answers to those of all political candidates.⁷ We only included two policy questions in the conjoint experiment to avoid the cognitive burden of overloading respondents with information.

⁵ We included a variety of the most popular traditional Danish-sounding names and the most popular immigrant-origin names used in Denmark. In total, 32 names were used.

⁶ Party-affiliation has proven to be the single most important information for voters (Green, Palmquist, and Schickler 2002).

⁷ Although some combinations of party affiliation and policy position are rare, all combinations are present in the actual pool of candidates.

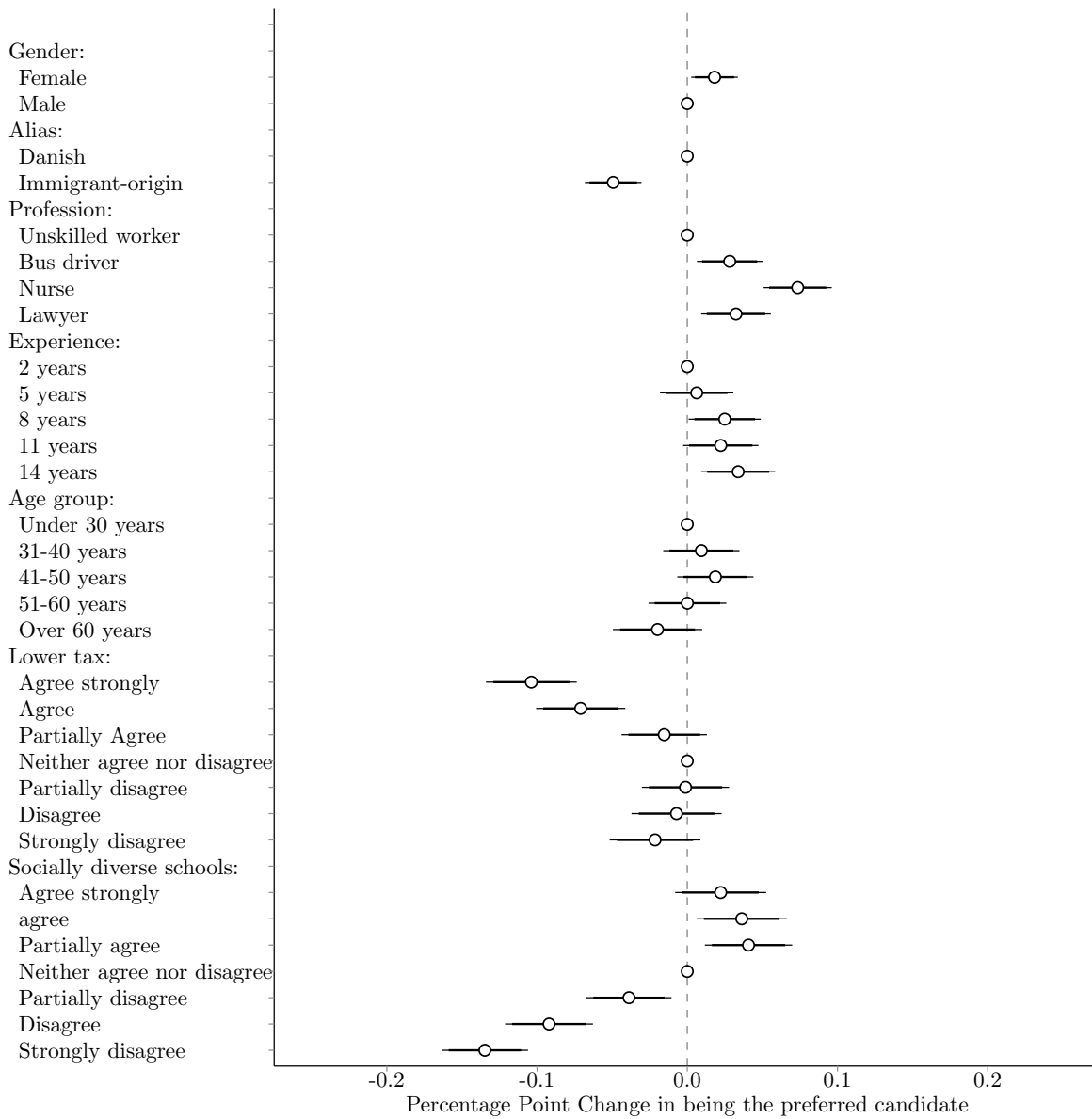
Results

As a first-order question, we explore whether Danish voters discriminate against immigrant-origin minority and female political candidates. Because the effects of each attribute level are measured on the same outcome, their relative importance can be evaluated by estimating Average Marginal Component Effects (AMCE). For example, we estimate the average difference in the probability of a candidate winning based on gender when the average for all possible combinations of the other attributes is computed. In other words, each estimate is the difference in the likelihood of being preferred as a candidate compared to one omitted attribute level (the reference category) when averaging over all other attributes. This is estimated by regressing an indicator for whether the respondent chooses a candidate on the specifications of the various characteristics mentioned above.⁸

Figure 3 depicts the AMCE of each attribute level and the 95% and 90% confidence intervals. The results indicate that potential political candidates with immigrant-origin names are significantly less preferred by voters. Specifically, voters on average prefer majority candidates over immigrant-origin candidates by five percentage points. Having an immigrant-origin name results in a larger (negative) estimate than the positive effect of having many years of political experience (as compared to no experience), being a lawyer (as compared to being unskilled) and roughly the same as being a nurse (as compared to being unskilled).

As seen from the figure, the results render little support for the notion that voters exhibit a direct bias against female candidates. If anything, voters seem to have a small, statistically insignificant preference for female over male candidates by 1.8 percentage points – a finding that aligns with a recent meta-analysis on voter preferences over gender (Schwarz, Hunt, and Coppock 2018). The figure also shows a number of other interesting results. For example, candidates with political experience are favoured by voters. However, unsurprisingly the effect sizes are biggest on the policy questions, where voters on average have a negative view of candidates who want to lower taxes and argue against socially diverse schools.

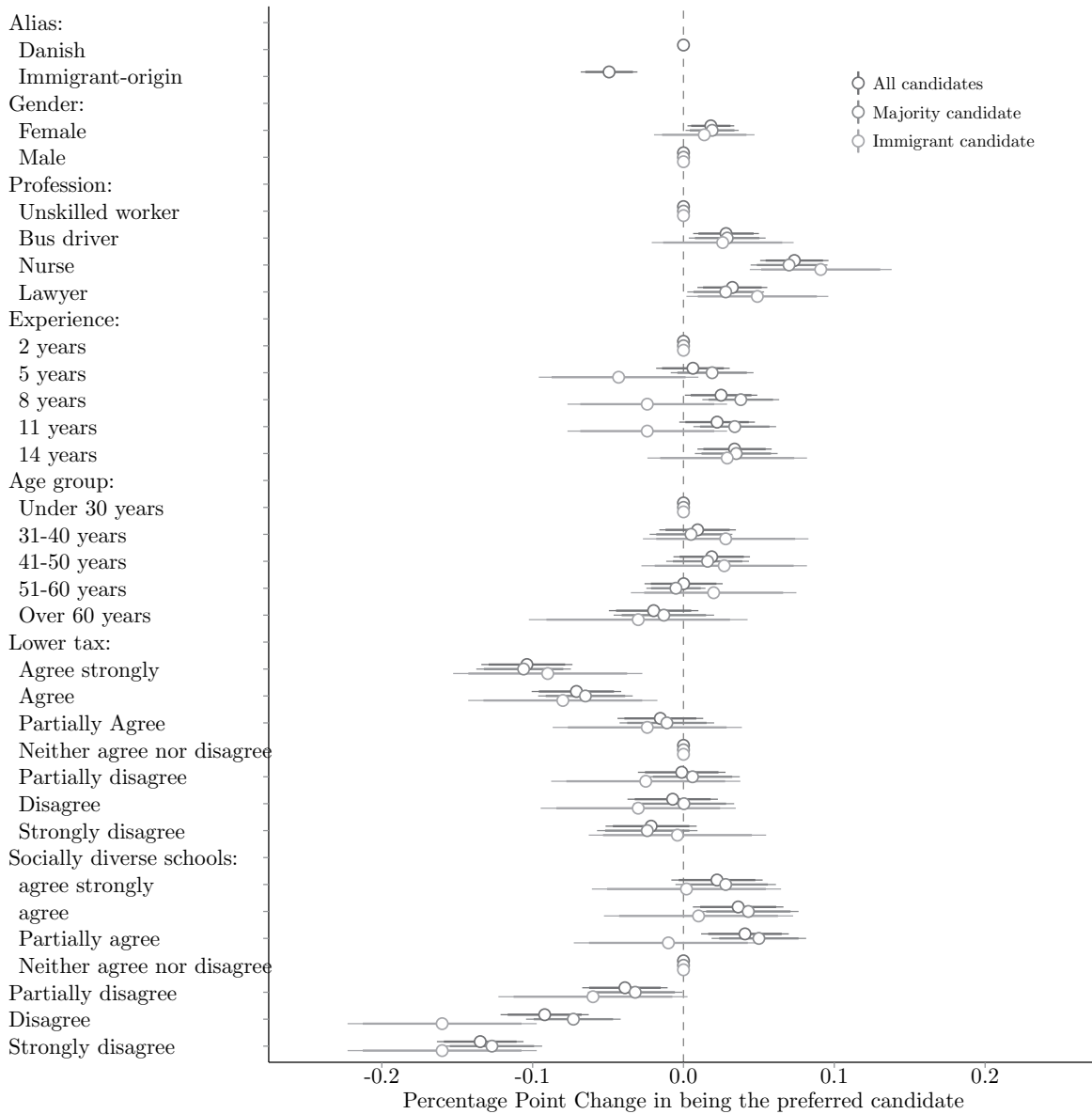
⁸ To correct for the within-respondent clustering, we cluster standard errors by respondent as is common in the literature (Hainmueller, Hopkins, and Yamamoto 2014).

Figure 3. Conjoint experiment: Effect of candidate attributes

Note: The figure indicates the average marginal component effect for each attribute level. 95% and 90% confidence intervals reported.

To examine if immigrant-origin minority candidates are punished or rewarded for different attributes than majority candidates, we repeat the analysis for majority and immigrant-origin candidates separately in the second and third row of Figure 4. Here, the data shows some heterogeneity in what voters reward depending on whether they evaluate an ethnic majority or immigrant-origin candidate. Immigrant-origin candidates appear to be rewarded less for political experience when compared to majority candidates. However, they are rewarded more for having a profession that requires an education (i.e. nurse or lawyer). This could indicate that information that contradicts stereotypes about immigrants – that they are less educated than average citizens – are positively rewarded. We also see that candidates who disagree or strongly disagree that municipalities should ensure social diversity at schools are evaluated more negatively when they have an immigrant-origin alias.

In the same vein, even though voters are less biased against female candidates overall, it may be that they are judged differently by voters than male candidates. As a result, they may be punished or rewarded for different attributes than men. We therefore repeat the analysis by exploring reactions to male and female candidates respectively in Appendix C. The general pattern is one of effect homogeneity: the effect of a candidate’s gender does not appear to depend much on the levels of other characteristics. Moreover, we also explore the actual distribution of the traits used in the conjoint experiments for real world candidates in Appendix D. We can see that the female candidates in general fit voter preferences better than male candidates. They are less likely to have an immigrant-origin name, are better educated, less likely to be over 60 years of age and have preferences on the tax and social diversity question, which align more with the policy preferences of the respondents in our survey. For example, a larger share of female candidates disagrees strongly with lowering taxes (46.1%) than male candidates (34.7%). However, female candidates are less likely to have been elected in the last election compared to male candidates (16.9% v. 20.5%). This means that female candidates may be more likely to face an incumbency disadvantage (Dahlggaard 2016).



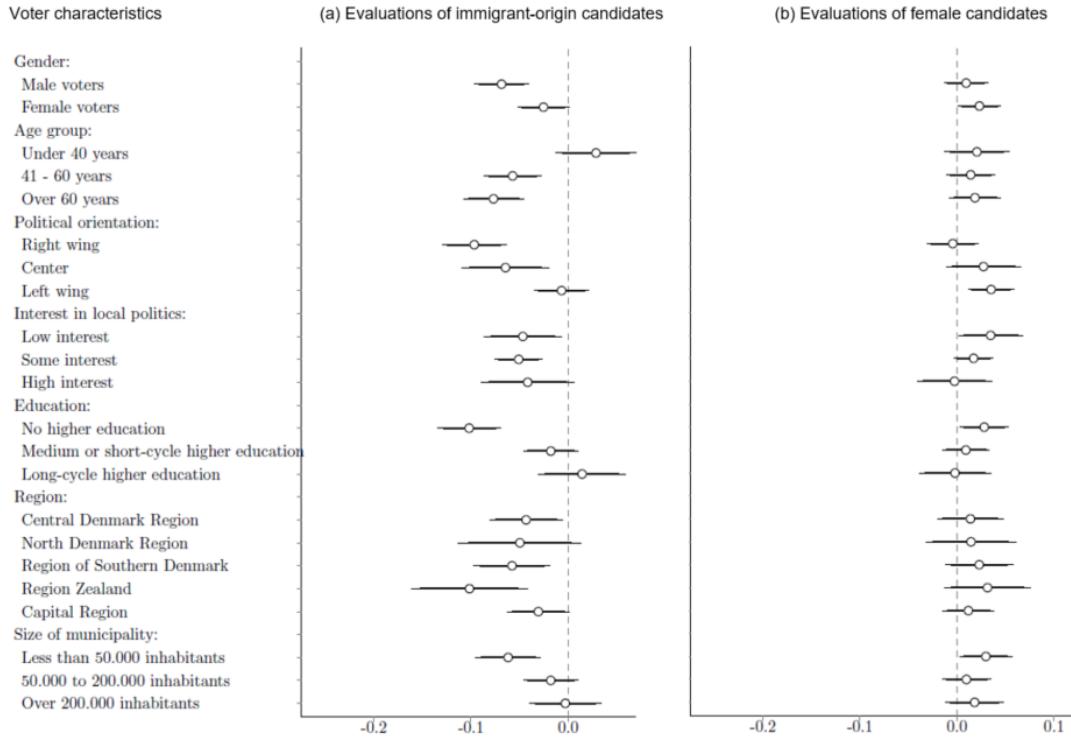
Note: The figure indicates the average marginal component effect for each attribute level. The second row shows the AMCE for majority candidates, while the third row shows the AMCE for candidates with an immigrant-origin alias.

Effect heterogeneity across voter characteristics

In this section we turn to the effects of candidates' ascriptive traits conditional on voter characteristics. This is important for the assessment of potential consequences of voter preferences. If the bias against immigrant-origin minority candidates is only prevalent among far right-wing party voters, the implications are arguably less surprising and less consequential for real-world immigrant-origin candidates who often run for left-wing parties. In the same vein, while there are small positive effects favouring female candidates in local elections, this does not mean that female candidates do not face a disadvantage among certain types of influential voters. For example, it may be the case that voters in small municipalities, where fewer votes are required to make it into office, have a stronger preference for male candidates than voters in large municipalities, or it may be that voters with specific political viewpoints systematically vote against certain candidates.

Figure 5 shows the average marginal component effects of having an immigrant-origin minority name (left) and a female name (right) across voter characteristics. The figure indicates a consistent preference for the ethnic majority candidate. Only among young and well-educated voters do we see a small preference in favour of the immigrant-origin candidate. If we look across the political spectrum, it is seen that there is essentially no effect among left-wing voters, whereas center- and right-wing voters have a strong preference for majority candidates.

Figure 5 indicates that female candidates are preferred by both male and female voters, although the difference is only significant for female voters. Likewise, all age groups prefer female candidates in the conjoint experiment. If we look across the political spectrum, it is clear that left-wing voters have a strong preference for female candidates, while right-wing voters are indifferent to the gender of the candidate. This is reflected in the actual composition of political parties and governing bodies, where left-wing parties have a higher share of elected female candidates than those on the right. Likewise, we see that voters with low educational achievement are those most supportive of female candidates. We see that female candidates are preferred across all regions and in both small and large municipalities.

Figure 5. Effects across voters' characteristics

Note: Effects across voters' characteristics of having (a) an immigrant-origin name compared to a majority name; and (b) being female compared to male.

Discussion and conclusion

This study set out to examine if voters condition their evaluations of local political candidates according to the candidates' gender and ethnicity. Our aim was to identify if these ascriptive characteristics matter for voters' preferences *even* when they have abundant information about the political candidates.

The results indicate a significant average voter preference for ethnic majority candidates over immigrant-origin candidates. We find little support for any differential treatment by candidates' gender – if anything, Danish voters overall have a small preference for female candidates which

aligns with a recent meta-analysis on voter preferences over gender (Schwarz, Hunt, and Coppock 2018). In addition, we find that the preference for female candidates is consistent across various respondent characteristics and we find no evidence of double standards either.

The preferential treatment based on candidates' ethnicity aligns with prior research that suggests that ascriptive characteristics influence voters' perceptions of candidates. While this is typically explained by the fact that these social categories serve as information shortcuts that allow voters to infer candidates' issue positions or party affiliation (Kirkland and Coppock 2017; McDermott 1998), our results indicate that even when such information is provided, voters evaluate political candidates based on salient ascriptive traits. We interpret this as evidence for the so-called 'electoral discrimination thesis' (Portmann and Stojanović 2019), according to which voters tend to discriminate against minority candidates. Unsurprisingly, the preference for ethnic majority candidates is larger among right-wing voters, but differential treatment is present across various respondent characteristics.

It is worth mentioning a number of limitations of our study. First and foremost, one concern is that the results reflect methodological limitations rather than actual voter preferences. Candidate choice conjoint experiments present generic descriptions of fictitious candidates rather than profiles of real-world political candidates. In other words, respondents in our conjoint experiment may react differently to hypothetical candidates than they would in the context of real elections. Another methodological objection is that respondents in an experimental setting may be aware that their responses are being evaluated (Berinsky 2004) and thus, respondents may reply in a way that is socially desirable. However, if anything this would lead us to underestimate the bias against immigrant-origin candidates. Moreover, the ability to mitigate social desirability bias is an often-mentioned advantage of conjoint experiments (Hainmueller, Hopkins, and Yamamoto 2014; Teele, Kalla, and Rosenbluth 2018; Sen 2017).

It is also worth considering how the traits that we included in the experiment could affect our results. The experiment included information on the hypothetical candidates' positions on two policy issues, which to some extent may crowd out the potential disadvantages of candidates background characteristics. Respondents' preferences regarding candidates may look different in

an experiment that provides less or different types of information (Pedersen, Dahlgaard, and Citi 2019). We leave it to future studies to advance this notion.

Regardless of these limitations, our study sheds light on ongoing debates both in political science and more broadly in discussions of the under-representation of immigrant-origin minorities in political bodies. The normative case for representation has been made in terms of its assumed effects on the nature and direction of public policy. Ample evidence indicates that politicians are more likely to advance the interests and preferences of groups that share their characteristics, including their profession, class, ethnicity and gender (Burden 2007; Carnes 2012; Grose, Mangum, and Martin 2007; Juenke and Preuhs 2012). Although incumbency status and position on the ballot is likely more determinative of electoral success (Portmann and Stojanović 2019), the fact that voters – across voter characteristics – are negatively biased towards candidates with certain ascriptive traits contributes directly to undermining the electoral prospects of these groups. Taking the point estimates from the experiment at face value, the electoral penalty appears to be considerable in magnitude.

References

- Atkeson, Lonna Rae and Brian T. Hamel (2018). “Fit for the Job: Candidate Qualifications and Vote Choice in Low Information Elections”. In: *Political Behavior*, pp. 1–24.
- Bækgaard, Martin and Ulrik Kjær (2012). “The gendered division of labor in assignments to political committees: Discrimination or self-selection in Danish local politics?” In: *Politics & Gender* 8.4, pp. 465–482.
- Banducci, Susan A et al. (2008). “Ballot photographs as cues in low-information elections”. In: *Political psychology* 29.6, pp. 903–917.
- Bauer, Nichole M (2015). “Emotional, sensitive, and unfit for office? Gender stereotype activation and support female candidates”. In: *Political Psychology* 36.6, pp. 691–708.
- Berinsky, Adam J. (2004). “Can we talk? Self-presentation and the survey response”. In: *Political Psychology* 25.4, pp. 643–659.
- Bernhard, Rachel and Sean Freeder (2018). “The More You Know: Voter Heuristics and the Information Search”. In: *Political Behavior*, pp. 1–21.
- Besco, Randy (2015). “Rainbow coalitions or inter-minority conflict? Racial affinity and diverse minority voters”. In: *Canadian Journal of Political Science/Revue canadienne de science politique* 48.2, pp. 305–328.
- Broockman, David and Evan Soltas (2017). *A natural experiment on taste-based racial and ethnic discrimination in elections*.
- Brouard, Sylvain and Vincent Tiberj (2010). *Yes they can: An experimental approach to eligibility of ethnic minority candidates in France*.
- Brown, Clyde, Neil R. Heighberger, and Peter A. Shocket (1993). “Gender-based differences in perceptions of male and female city council candidates”. In: *Women & Politics* 13.1, pp. 1–17.
- Burden, Barry C. (2007). *Personal roots of representation*. Princeton University Press.
- Campbell, Rosie and Philip Cowley (2014). “What voters want: Reactions to candidate characteristics in a survey experiment”. In: *Political Studies* 62.4, pp. 745–765.

- Carnes, Nicholas (2012). “Does the numerical underrepresentation of the working class in Congress matter?” In: *Legislative Studies Quarterly* 37.1, pp. 5–34.
- Carnes, Nicholas and Noam Lupu (2016). “Do voters dislike working-class candidates? Voter biases and the descriptive underrepresentation of the working class”. In: *American Political Science Review* 110.4, pp. 832–844.
- Carroll, Susan J. and Kira Sanbonmatsu (2013). *More women can run: Gender and pathways to the state legislatures*. Oxford University Press.
- Cutler, Fred (2002). “The simplest shortcut of all: Sociodemographic characteristics and electoral choice”. In: *Journal of Politics* 64.2, pp. 466–490.
- Dahlgaard, Jens Olav (2016). “You just made it: Individual incumbency advantage under Proportional Representation”. In: *Electoral Studies* 44, pp. 319–328.
- Dancygier, Rafaela M. et al. (2015). “Why are immigrants underrepresented in politics? Evidence from Sweden”. In: *American Political Science Review* 109.4, pp. 703–724.
- Dolan, Kathleen (2010). “The impact of gender stereotyped evaluations on support for women candidates”. In: *Political Behavior* 32.1, pp. 69–88.
- (2014). “Gender stereotypes, candidate evaluations, and voting for women candidates: what really matters?” In: *Political Research Quarterly* 67.1, pp. 96–107.
- Eurostat (2017). *Total general government expenditure*. data retrieved from Eurostat, <http://ec.europa.eu/eurostat/tgm/refreshTableAction.do?tab=table&plugin=1&pcode=tec00023&language=en>.
- Fisher, Stephen D et al. (2015). “Candidate ethnicity and vote choice in Britain”. In: *British Journal of Political Science* 45.4, pp. 883–905.
- Fox, Richard L. and Jennifer L. Lawless (2010). *It still takes a candidate: Why women don’t run for office*. Cambridge University Press.
- Green, Donald, Bradley Palmquist, and Eric Schickler (2002). *Partisan Hearts and Minds*.
- Grose, Christian R., Maruice Mangum, and Christopher Martin (2007). “Race, political empowerment, and constituency service: Descriptive representation and the hiring of African-American congressional staff”. In: *Polity* 39.4, pp. 449–478.

- Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto (2014). "Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments". In: *Political Analysis* 22.1, pp. 1–30.
- Huddy, Leonie and Nayda Terkildsen (1993). "The consequences of gender stereotypes for women candidates at different levels and types of office". In: *Political Research Quarterly* 46.3, pp. 503–525.
- Juenke, Eric Gonzales and Robert R. Preuhs (2012). "Irreplaceable Legislators". In: *Rethinking Minority*.
- Kage, Rieko, Rosenbluth, and Seiki Tanaka (2017). "What Explains Low Female Political Representation? Evidence from Survey Experiments in Japan". In:
- Kirkland, Patricia A. and Alexander Coppock (2017). "Candidate Choice Without Party Labels". In: *Political Behavior*, pp. 1–21.
- Kjær, Ulrik (2013). "Efter de første 100 år – kvinder i dansk kommunalpolitik 1909-2009". In: *KV09. Analyser af kommunalvalget 2009*. Ed. by Jørgen Elklit and Ulrik Kjær. Odense: Syddansk Universitetsforlag. Chap. 17, pp. 321–338.
- Lawless, Jennifer L. (2015). "Female candidates and legislators". In: *Annual Review of Political Science* 18, pp. 349–366.
- Lawless, Jennifer L. and Kathryn Pearson (2008). "The primary reason for women's underrepresentation? Reevaluating the conventional wisdom". In: *The Journal of Politics* 70.1, pp. 67–82.
- McConaughy, Corrine M et al. (2010). "A Latino on the ballot: Explaining coethnic voting among Latinos and the response of White Americans". In: *The Journal of Politics* 72.4, pp. 1199–1211.
- McDermott, Monika L. (1998). "Race and gender cues in low-information elections". In: *Political Research Quarterly* 51.4, pp. 895–918.
- McElroy, Gail and Michael Marsh (2010). "Candidate gender and voter choice: Analysis from a multimember preferential voting system". In: *Political Research Quarterly* 63.4, pp. 822–833.
- Mo, Cecilia Hyunjung (2015). "The consequences of explicit and implicit gender attitudes and candidate quality in the calculations of voters". In: *Political Behavior* 37.2, pp. 357–395.

- Norris, Pippa and Joni Lovenduski (1995). *Political recruitment: Gender, race and class in the British Parliament*. Cambridge University Press.
- Ono, Yoshikuni and Masahiro Yamada (2016). “Do voters prefer gender stereotypic candidates? evidence from a conjoint survey experiment in Japan”. In: *Political Science Research and Methods*, pp. 1–16.
- Pedersen, Rasmus T, Jens Olav Dahlgaard, and Manuele Citi (2019). “Voter reactions to candidate background characteristics depend on candidate policy positions”. In: *Electoral Studies* 61, p. 102066.
- Philpot, Tasha S and Hanes Walton Jr (2007). “One of our own: Black female candidates and the voters who support them”. In: *American Journal of Political Science* 51.1, pp. 49–62.
- Portmann, Lea and Nenad Stojanović (2019). “Electoral discrimination against immigrant-origin candidates”. In: *Political Behavior* 41.1, pp. 105–134.
- Sanbonmatsu, Kira (2002). “Gender stereotypes and vote choice”. In: *american Journal of political Science*, pp. 20–34.
- Schwarz, Susanne, William Hunt, and Alexander Coppock (2018). “What Have We Learned About Gender From Candidate Choice Experiments? A Meta-analysis of 30 Factorial Survey Experiments”. In: The working paper is found at https://alexandercoppock.com/papers/SHC_gender.pdf.
- Sen, Maya (2017). “How political signals affect public support for judicial nominations: Evidence from a conjoint experiment”. In: *Political Research Quarterly* 70.2, pp. 374–393.
- Shames, Shauna L. (2017). *Out of the running: Why millennials reject political careers and why it matters*. NYU Press.
- Stegmaier, Mary, Michael S Lewis-Beck, and Kaat Smets (2012). “Standing for Parliament: Do Black, Asian and Minority Ethnic Candidates Pay Extra?” In: *Parliamentary affairs* 66.2, pp. 268–285.
- Street, Alex (2014). “Representation despite discrimination: minority candidates in Germany”. In: *Political Research Quarterly* 67.2, pp. 374–385.

- Teele, Dawn L., Joshua Kalla, and Rosenbluth (2018). “The Ties that Double Bind: Social Roles and Women’s Underrepresentation in Politics”. In: *American Political Science Review*, pp. 1–17.
- Thrasher, Michael et al. (2017). “Candidate ethnic origins and voter preferences: Examining name discrimination in local elections in Britain”. In: *British Journal of Political Science* 47.2, pp. 413–435.

**Supplementary material for Candidate choice in a high-information
setting: Do ascriptive characteristics shape candidates' electoral
prospects?**

Appendix A. Sample Characteristics

	Sample	Population
Female	49.7%	50.6%
Age, mean (sd)	53 (16.6)	49.1 (18)
Education (share with tertiary-level education)	49%	34.1%
Left-right position, 0-1 (sd)	.48 (.24)%	n.a.%

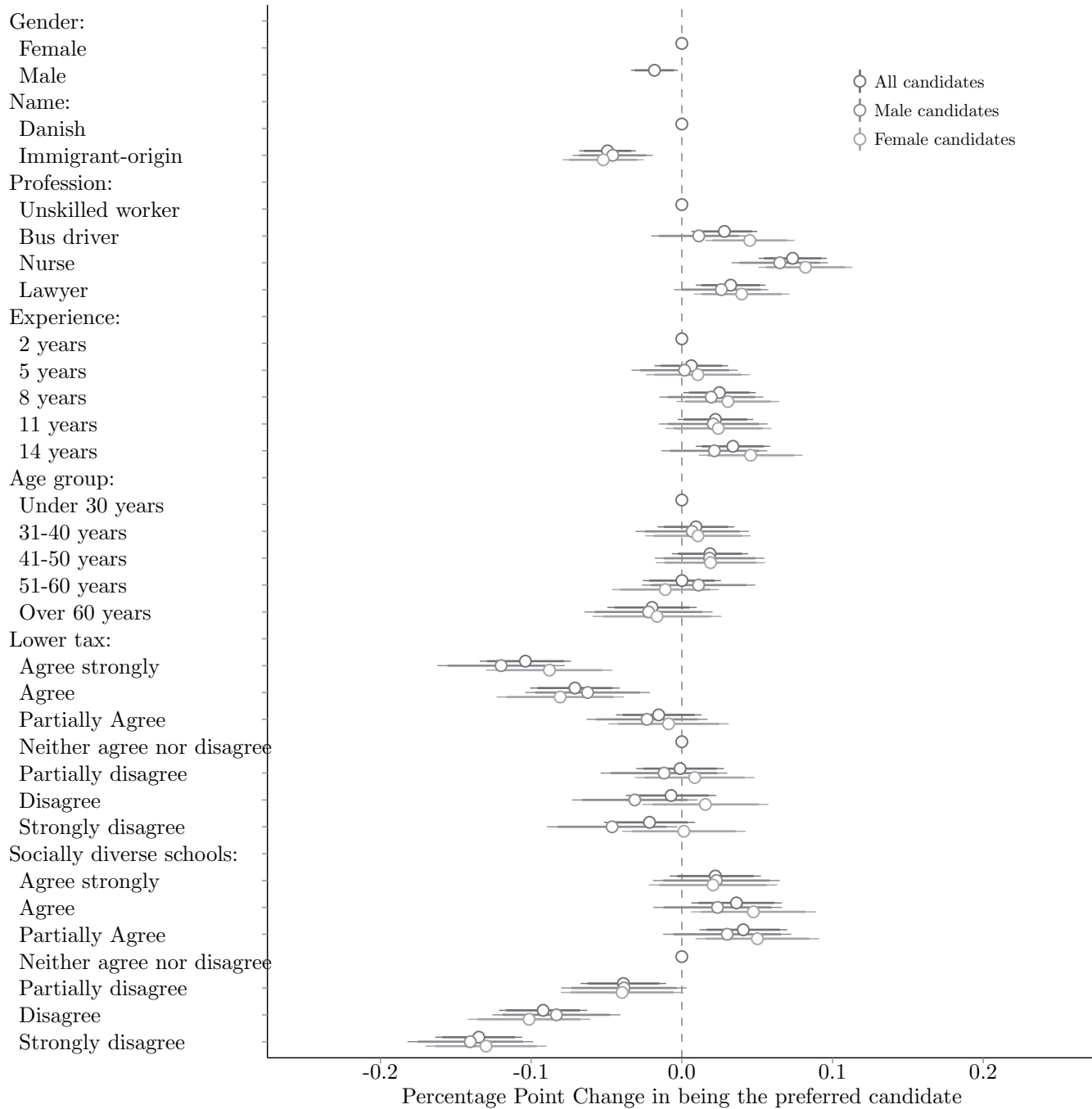
Note: Population data are for Danish citizens aged > 18. Source: Statistics Denmark (www.statistikbanken.dk)

Appendix B. Example of conjoint pair

<p>Suppose that the following two potential candidates run for a seat in the municipality council. Which of the two candidates would you like to vote for? Even if you are not entirely sure, please indicate which of the two you would prefer if you had to choose either one of them.</p>		
Candidate A		Candidate B
55	Age	40
Peter	Name	Anne
Unskilled	Educational background	Lawyer
7 years	Political experience	11 years
Strongly agree	Should local taxes be reduced?	Agree
Agree	Should municipalities secure diversity in schools?	Disagree
Liberal Party	Party affiliation	Liberal Party

Appendix C.

In this appendix, we repeat the analysis for male and female candidates separately in the second and third row. Here, it is seen that the results in general are very similar for male and female candidates. Nonetheless, there are small, but insignificant differences. Female candidates are rewarded more for having a skilled job and for having more experience in local politics relative to men. Most interestingly, women are punished less for having extreme views in either directions on whether taxes should be lowered. The pattern is less clear for the question on socially diverse schools. This could either indicate that voters are more tolerant towards female candidates with extreme views than male candidates or that female candidate's views on tax policies are less important to voters when they make their decision. This fluctuate well with the research on gender stereotypes, which has shown that male politicians are considered to be more competent on "hard" policy questions than female politicians (Huddy and Terkildsen 1993).

Figure C. Effect of candidate attributes broken down by candidate gender

Note: The figure indicates the average marginal component effect for each attribute level. The second row shows the AMCE for male candidates, while the third row shows the AMCE for female candidates.

Appendix D. The distribution of traits for the real pool of candidates

The table below shows the distribution of traits from the conjoint experiment for the real world candidates. All the data are from the latest election in 2017 apart from the data on the candidates' education, which is only available for the election in 2013.

	Women	Men	Total	Year
Name				
Immigrant-origin	2.3%	3.4%	3.1%	2017
Danish	97.7%	96.6%	96.9%	2017
Profession				
No higher education	12.5%	15.7%	14.7%	2013
Vocational education	25.4%	37.6%	33.9%	2013
Medium or short-cycle higher education	37.8%	24.9%	28.8%	2013
Long-cycle higher education	24.3%	21.8%	22.6%	2013
Experience				
Elected in last election	16.9%	20.5%	19.4%	2017
Not elected in last election	83.1%	79.5%	80.6%	2017
Age group:				
Under 30 years	8.4%	8.3%	8.4%	2017
30-39 years	12.9%	10.4%	11.2%	2017
40-49 years	25.7%	20.8%	22.3%	2017
50-59 years	26.5%	26.2%	26.3%	2017
Over 60 years	26.6%	34.3%	31.9%	2017
Lower tax:				
Disagree strongly	46.1%	34.7%	38.3%	2017
Partially disagree	13.4%	14.9%	14.4%	2017
Neither agree nor disagree	13.6%	15.2%	14.7%	2017
Partially agree	11.5%	14.8%	13.8%	2017
Agree strongly	15.4%	20.5%	18.9%	2017
Socially diverse schools:				
Disagree strongly	6.6%	8.0%	7.6%	2017
Partially disagree	9.4%	9.7%	9.6%	2017
Neither agree nor disagree	11.3%	10.0%	10.4%	2017
Partially agree	33.9%	31.6%	32.3%	2017
Agree strongly	38.7%	40.7%	40.1%	2017

Social desirability bias in conjoint experiments: What is the optimal design when studying sensitive topics?

Malte Dahl*

Under review in Political Analysis

Abstract

An often-mentioned advantage of conjoint experiments over traditional survey experimental designs is that the former have the potential to mitigate social desirability bias. To what extent this is true may depend on a number of design choices – a concern that has received surprisingly little empirical attention. I conducted two studies in which I randomly assigned respondents to three types of conjoint designs in order to manipulate their awareness to sensitive features and possibilities for justifying inappropriate answers ($N = 7,059$). The results show that design variations significantly affect respondents' inferences about the research objective. However, there are no detectable differences between respondents' preferences across designs. This indicates that researchers using conjoint experiments should not compromise their choice of design because of concerns over social desirability bias.

*I would like to thank Mogens Jin Pedersen, Alex Coppock, Thad Dunning, Peter T. Dinesen, Frederik Hjørth, Kasper Moeller Hansen, Anders Woller, Ma Yi, Benjamin Egerod, Jens van der Ploek and participants at the seminar *New approaches for the study of political behaviour* at the Department of Political Science, University of Copenhagen. Also, a big thanks to Thomas Leeper for creating the R package *cregg*.

Introduction

Conjoint experiments have become a standard part of the political science toolkit. These experiments are effective and low-cost tools that enable researchers to elucidate respondents' multidimensional preferences and test several causal hypotheses simultaneously (Hainmueller, Hopkins, and Yamamoto 2014; Hainmueller, Hangartner, and Yamamoto 2015; Bansak, Hainmueller, Hopkins, et al. 2017). Moreover, researchers can easily increase the effective sample size by letting each individual respondent answer several conjoint.¹

Another considerable advantage that is often emphasised by proponents of conjoint experiments is that these designs have the potential to mitigate social desirability bias (SDB) (Hainmueller, Hopkins, and Yamamoto 2014; Horiuchi, Smith, and Yamamoto 2017; Teele, Kalla, and Rosenbluth 2018). The ability to obtain reliable answers is a key inferential issue in the survey-experimental literature and considering that conjoint experiments are often used to gauge respondents' reactions to sensitive dimensions, this is an essential quality. However, despite the prominence of conjoint designs, there has been surprisingly little effort to examine the extent to which, and the conditions under which, SDB is of concern when examining sensitive topics.

The perceived ability of conjoint experimental designs to mitigate SDB is grounded in two notions. First, since respondents are presented with numerous features, a given sensitive feature is 'masked' among other features that are also randomly varied (*an attention assumption*). Therefore, it is argued, respondents cannot infer that the sensitive feature is of particular importance (Teale, Kalla, and Rosenbluth 2018). Second, respondents can always find multiple justifications for any given choice (Hainmueller, Hopkins, and Yamamoto 2014). This implies that inappropriate answers can be justified by (combinations of) the levels of other features in the experiment (*a justification assumption*).

The extent to which these two assumptions hold may be heavily conditioned by a number of specific design choices. For example, there is a fundamental difference between paired and

¹ Conjoint experiments have been used to study how voter preferences are shaped by political candidates' gender (Teale, Kalla, and Rosenbluth 2018) and class (Carnes and Lupu 2016), the way information on party affiliation moderates voter preferences (Kirkland and Coppock 2017), and Americans' attitudes towards immigrants (Hainmueller and Hopkins 2015).

single-profile designs, between designs that measure outcomes as a discrete choice, a rating or a combination of the two, and between designs that manipulate few or many features and feature levels.² Moreover, different randomisation schemes can be applied, with some studies randomising the number of features that are presented, randomising all or only some of the features and/or randomising feature levels with different probability weights (Hainmueller and Hopkins 2015). I argue, that these design differences are likely to have an effect on social desirability pressures because they influence (i) respondents' anticipation of the primary research objective, and (ii) the degree to which respondents can justify inappropriate answers over repeated tasks. I also argue that conjoint designs that, at least in theory, downplay social desirability pressures often compromise other important features of the experiment (e.g. statistical power or ecological validity). This raises an important question: what is the optimal design when studying sensitive topics in conjoint experiments?

In this pre-registered study,³ I aim to answer that question by randomly assigning respondents to seemingly similar conjoint designs that vary social desirability pressures. To do so, I ran two studies inspired by Sen (2017) and Hainmueller and Hopkins (2015), respectively. Both of these were conducted using Amazon's Mechanical Turk marketplace ($N = 7,059$). In each study, respondents were randomly assigned to one of three conjoint designs intended to either minimise or amplify their attention to sensitive dimensions and their possibilities for justifying inappropriate answers.⁴ Specifically, in the first condition, *the high-contrast paired design*, each respondent was presented with a number of conjoint pairs in which the levels of a sensitive feature were repeatedly contrasted (e.g., a black vs. a white candidate). The second condition, *the restricted paired design*, was similar, except that the sensitive feature was only contrasted in a limited number of conjoint pairs. Finally, the third condition was a fully randomised *single-profile design* showing only one candidate at a

² *Features* can include, for example, age, party affiliation and gender, whereas *feature levels* are the values each feature can take, e.g., male/female in the case of gender.

³ The project was registered at Open Science Framework and a pre-analysis plan of Study 1 can be found at www.osf.io/sf6h9, while a pre-analysis plan for study 2 can be found at www.osf.io/ket62

⁴ This work is related to recent studies that have examined demand effects in survey experiments by inducing different degrees of information about the purpose of the study (Mummolo and Peterson 2018; De Quidt, Haushofer, and Roth 2017). However, instead of raising awareness of the research objective by providing respondents with explicit information, the present project sought to manipulate awareness to sensitive dimensions *through design*.

time.

The results demonstrate that these design differences significantly affect respondents' inferences about the research objective (i.e. their attention to a sensitive feature). Specifically, respondents assigned to a high-contrast paired conjoint design are much more likely to infer that the sensitive feature is the main focus of the study compared to respondents assigned to either of the other two designs. Surprisingly, and most importantly, the design differences do not translate into any immediate effect on respondents' priorities. When comparing the effects of the sensitive features across designs, there are no distinguishable differences: respondents' answers are stable. This evidence indicates that when researchers use conjoint designs to study sensitive topics, they should not compromise their choice of design due to the fear of SDB.

Social desirability bias in survey research

A common understanding of SDB is that it stems from respondents' lack of comfort to reveal their true attitudes (Tourangeau and Yan 2007; Groves et al. 2011; Holtgraves 2004). Respondents moderate their behaviour by giving normatively positive responses in order to make themselves look more favourably and avoid the embarrassment, unease and distress that revealing socially undesirable answers may bring (Kaminska and Foulsham 2013). For example, respondents tend to underreport favoritism for preferred groups relative to nonpreferred ones (Janus 2010; Kuklinski et al. 1997) which leads to a misrepresentation of preferences.

Evidence on SDB in survey research generally suggests that it is a valid concern. This is demonstrated in studies that word questions in more or less threatening ways (Kuklinski et al. 1997), that change the interview setting (Krysan and Couper 2003), that compare results from list experiments with direct questions (Janus 2010; Gilens, Sniderman, and Kuklinski 1998) or studies that compare survey answers with register data (Hariri and Lassen 2017).

Moreover, several studies have demonstrated that some groups of respondents are more likely to provide socially desirable answers. For example, Berinsky and Lavine (2012) demonstrate that high self-monitors are more likely to offer socially acceptable answers. Other studies indicate that liberal

respondents are more likely to give untruthful answers to questions regarding race (Gilens, Sniderman, and Kuklinski 1998) and immigration restrictionist policy questions (Janus 2010). A related concern is that survey experiments frequently rely on online subject pools, like Amazon’s Mechanical Turk, where experienced experimental participants have incentives to be especially attentive to researcher expectations (Krupnikov and Levine 2014). For this reason, Berinsky, Huber, and Lenz (2012) recommend that researchers avoid revealing their intentions in online survey experiments.⁵

Conjoint experiments as a means to overcome SDB

While SDB is a potential validity issue in all survey research, it is often claimed that conjoint experiments can mitigate some of these concerns (Hainmueller, Hopkins, and Yamamoto 2014; Liu 2018; Teele, Kalla, and Rosenbluth 2018). Two arguments support this idea. First, because respondents in conjoint experiments are typically presented with a large number of features, the design allows respondents to justify any particular choice or rating (Hainmueller, Hopkins, and Yamamoto 2014). Secondly, due to the large number of varying features, the main research objective of the study is unclear to respondents (Hainmueller, Hopkins, and Yamamoto 2014; Ono and Yamada 2016). For example, in a study of gender biases in voters’ evaluations of political candidates, Teele, Kalla, and Rosenbluth (2018) state that because candidate gender is embedded as one of multiple features ‘(...) *our own interest in gender would not have been obvious in the experiment. This likely lessens the degree to which our results are skewed by social desirability bias*’.

The notion that researchers can mitigate SDB and obtain more reliable answers when research intentions are ‘masked’ is not new. Previous survey research on sensitive topics have implemented cover stories in order to misdirect participants about the goal of the experiment (McDermott 2002; Dickson 2011). For example, by asking questions unrelated to the primary intention of the study (Kam 2007) or by providing respondents with an alternative or vaguely stated purpose of the experiment (Bullock 2011; Arceneaux 2008).

The arguments for why conjoint designs should minimize concerns over SDB appear plausible,

⁵ A researcher demand effect is distinct from SDB and happens when respondents infer the response researchers expect and behave in line with these expectations (Mummolo and Peterson 2018). In principle, demand effects could work in the opposite direction of SDB which I test in the final part of the paper.

but there is little empirical evidence to support them. On the one hand, some studies that use conjoint designs have implemented various tests in order to reject that SDB is an issue. For example, Bansak, Hainmueller, and Hangartner (2016) find that results are stable for respondents with different levels of empathy, building on the idea that empathy and social desirability scales correlate. Hainmueller and Hopkins (2015) come to the same conclusion after re-estimating their results based on measures of self-monitoring that are known to be closely connected to social desirability. Finally, Hainmueller, Hangartner, and Yamamoto (2015) use a natural experiment as a behavioural benchmark and compare the results from conjoint experiments with real-world behaviour.

On the other hand, results from several conjoint experiments that study sensitive dimensions seem at odds with what we know from field experiments or observational studies and run counter to observed real-world outcomes. For example, a number of studies on voter preferences that use candidate conjoint designs find no effects – or even positive effects – of being a non-white political candidate compared to a white political candidate (Carnes and Lupu 2016; Kirkland and Coppock 2017).⁶ These results contradict studies of actual voting patterns (e.g. Broockman and Soltas (2017) and Lewis-Beck, Tien, and Nadeau (2010)). This seems to indicate that the results from conjoint experiments may be biased because of SDB. This concern is further strengthened by recent evidence suggesting that experimental findings on voter preferences for women or black candidates may overestimate support, even in anonymous settings (Krupnikov, Piston, and Bauer 2016).

Finally, while Hainmueller, Hangartner, and Yamamoto (2015) demonstrate that the paired-conjoint design is aligned with real-world behaviour, they also demonstrate that *'seemingly subtle differences in survey designs can produce significant differences in performance'*. In summary, there is reason to suspect that SDB can be an issue in conjoint experiments, making it pertinent to understand if design adjustments can mitigate this type of response bias.

⁶ Carnes and Lupu (2016) conducted a conjoint experiment in which they manipulated candidates' race using two levels (white and black) in a study of support for political candidates, and find a positive (although only borderline significant) effect of being black. Similarly, Kirkland and Coppock (2017) finds that Hispanic, Black and Asian candidates respectively are preferred over White candidates (although these differences are not significant).

Research design

I conducted two independent studies each comprising three conjoint experiments specifically designed to assess the relation between design and SDB. The experiments are almost identical to two previous studies by Sen (2017) and Hainmueller and Hopkins (2015).⁷ The experiments were implemented in Qualtrics software and fielded in August 2018 on a total of 7,059 respondents recruited from Amazon’s Mechanical Turk, which hosts an experienced pool of survey respondents (Berinsky, Huber, and Lenz 2012).⁸

Manipulating attention to sensitive features through design

Both studies include a feature that is known to be influenced by social desirability pressures. Study 1 seeks to gauge the effect of candidates’ race, a topic to which it can be difficult to obtain honest self-reports since racial preferences is taboo (Krupnikov, Piston, and Bauer 2016; Berinsky and Lavine 2012). Study 2 seeks to explore support for immigrants seeking admission to the US. Religious affiliation, more specifically being Muslim, serves as a sensitive feature level. Restrictionist immigration policies is a hot-button topic that previous research has found to be subject to SDB (Janus 2010).

We can think of the identification strategy as a two-stage process. The first stage concerns the link between the specific design and respondents’ attention to sensitive features and their possibility of justifying inappropriate answers. The second stage concerns whether this affects respondents’ priorities. I seek to manipulate respondents’ awareness to the sensitive feature in two ways. First, I manipulate the probability weights of the levels of the sensitive feature across conditions. Thus, one condition, the *high-contrast design*, is a paired-conjoint in which respondents are presented with five different candidate pairs with each or most pairs displaying a contrast on the sensitive feature

⁷ The designs in the present study differ slightly from the original studies in terms of the number and type of features included. Considering that the purpose of the present study being not to replicate these studies, but rather to determine whether treatment effects vary across design, this is not problematic.

⁸ In 2018, researchers raised concerns that an increasing number of “bots” (respondents using semi- or fully-automated code to automatically respond) reduced the quality of answers to surveys fielded on Amazon’s Mechanical Turk. In order to weed out potential bots I used reCAPTCHA and a basic quality check (What is 2+2?).

(for example, a black vs. a white candidate). Arguably, the repeated contrast increase respondents' awareness to the sensitive feature. Moreover, the frequent contrast makes it harder for respondents to defend an inappropriate answer since they have to repeat it across five conjoint pairs. We would expect SDB to amplify in this condition. In the second condition, the *restricted paired design*, the sensitive feature is contrasted less frequently. Thus, the restriction serves to mask the sensitive feature from respondents by design.

Secondly, I test the importance of the within-subject structure that characterizes the paired design by including a *single-profile design* as a third condition (See details on the conjoints in appendix A). While respondents in a paired conjoint design observe *both* treatment and control at the same time, the single-profile conjoint displays *either* control or treatment which arguably makes the sensitive feature less noticeable. Again, I expect this design to reduce social desirability pressures compared to the high-contrast paired design.

In each study, respondents are randomly assigned to one of the three conditions. Because the second condition is restricted on the sensitive feature which reduces statistical power, half of the respondents are assigned to this condition in order to gain precision, while a quarter of the sample is assigned to the high-contrast design and the single-profile design respectively.

Study 1: U.S. Supreme Court nominees

The first study is inspired by a candidate conjoint study on support for Supreme Court nominees by Sen (2017). The design is a typical example of a conjoint design in which the researcher asks a sample of 1,650 U.S. adults to rank a number of hypothetical candidates. While the original study used three different outcome measures on a 7-point likert scale ("Support", "Qualifications", and "Trust"), I only ask respondents to either indicate who they support most or, in the single-profile, to rate their level of support for the candidate. Also, I exclude information on political leaning that was assigned to half of the respondents in the original study.

Respondents are randomly assigned to one of the three conjoint experiments that are otherwise identical in terms of features, levels, wording and formatting. The experiments include six features that each hold several feature levels (See details in appendix B). Most importantly, candidates' race

are assigned from a list with two levels (black or white).

In the high-contrast condition respondents are presented with five different pairs in which each pair contrasts candidates' race. That is, all five candidate pairs appear as *Black vs. White* or *White vs. Black*. The second condition is equivalent to the first except candidates' race is restricted to appear only in one of the five pairs. In the final condition, respondents are presented with a single-profile conjoint in order to eliminate the contrast on race that is inherent to the comparison in paired-conjoint designs. In this condition the candidates' race is assigned randomly. The design is summarized in Figure 1.

Study 2: Immigrants seeking admission to the U.S.

Study 2 is substantively inspired by Hainmueller and Hopkins (2015) and examines respondents' support for immigrants applying for admission to the U.S.⁹ As in Study 1, respondents are assigned to one of three variations of a conjoint design that all include seven features (See details in appendix B). Most importantly here is *Religion* that can take on six levels (Catholic, Protestant, Jewish, Muslim, Atheist or Other). I follow roughly the same strategy as in study 1, and assign respondents to three different conjoint designs varying the focus on the sensitive feature. In the high-contrast design, the probability that one of the two profiles in any given pair is Muslim is high (80 per cent of all pairs), whereas in the second condition the probability that one of the two profiles is Muslim is restricted (17 per cent of all pairs). Finally, in a single-profile conjoint, religious affiliation is drawn randomly, but as was the case in study 1, the religious contrast is arguably not as prominent due to the non-paired structure of the design.

⁹ In the original study, the features were chosen to approximate the information available to immigration officials which is why religion was omitted, but the authors suggest religion as a dimension for future work to explore.

Figure 1. Experimental conditions in Study 1 and Study 2

	Paired conjoint, High contrast	Paired conjoint, Restricted	Single-profile conjoint
Study 1			
Feature: Race Levels: <i>Black / White</i>	5 of 5 pairs contrast race N = 854	1 out of 5 pairs contrasts race N = 1765	Random assignment of race N = 874
Study 2			
Feature: Religion Levels: <i>Muslim / Protestant / Catholic / Jewish / Atheist / Other</i>	80 % chance that one candidate is Muslim N = 926	17 % chance that one candidate is Muslim N = 1770	Random assignment of religion N = 870

Results

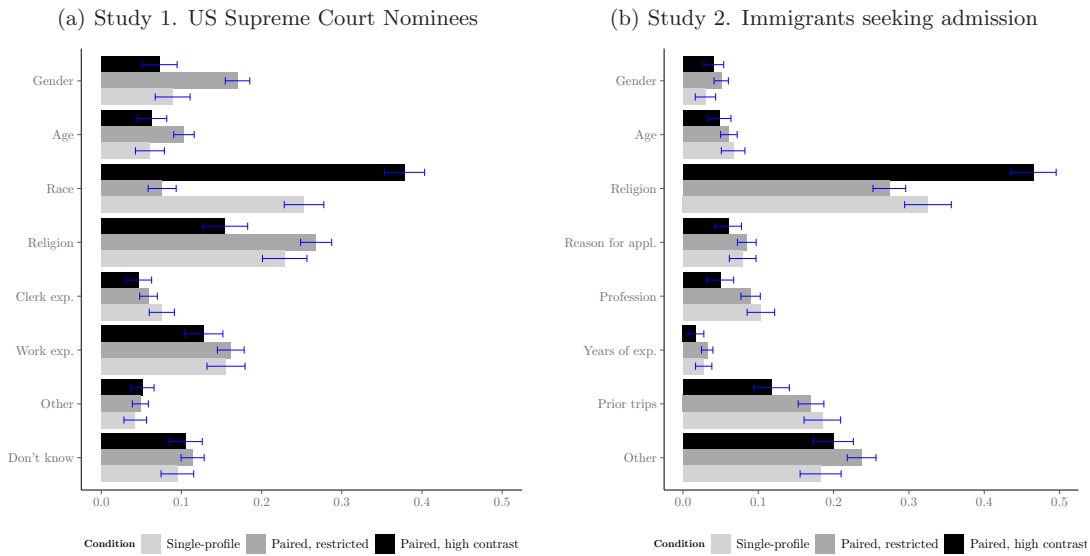
Can respondents infer research intentions?

A first-order concern is whether the design variations in fact have an effect on respondents' awareness to the sensitive feature. To check if this is the case, the survey included a post-treatment question asking respondents to choose from a list of eight different options what they believed to be the main objective of the study. As displayed in Figure 2 respondents' anticipation of the research objective changes drastically across design conditions. In Study 1, 38 percent of respondents in the high-contrast condition answered that the primary intent of the study was to examine their reactions to candidates' race. This is 30 percentage points more compared to the restricted paired conjoint, and 13 percentage points more relative to the single-profile conjoint. We see the same pattern in Study 2. 48 percent of respondents in the high-contrast paired design believed that the main objective of the study was to examine support for immigrants conditional on their religious affiliation, which is 20 percentage points more than in the restricted paired conjoint and 16 percentage points more compared to the single-profile conjoint.

This demonstrates two important points. First, that respondents generally pay much attention to sensitive features such as race or religious affiliation when answering these experiments. Second, that seemingly subtle design differences significantly affect respondents' inferences about research

intentions. In other words, it *is* possible to downplay a sensitive dimension by adjusting the design and thus make respondents significantly less likely to infer that the sensitive feature is important.

Figure 2. Manipulation check. Respondents' perception of the main research objective



Note: The figures indicate the distribution of respondents' anticipation of the main research objective across the three designs. Figure (a) at the left depicts the results from study 1, while figure (b) at the right shows the results from study 2.

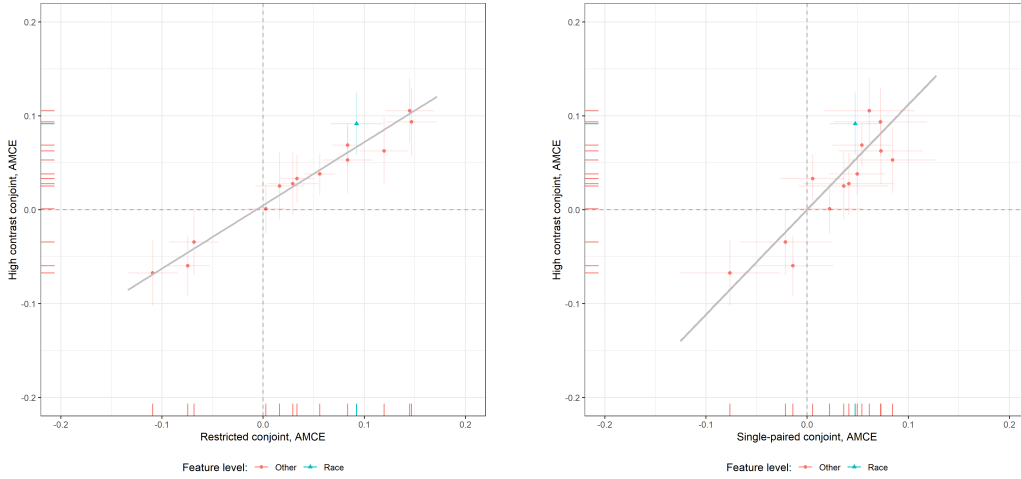
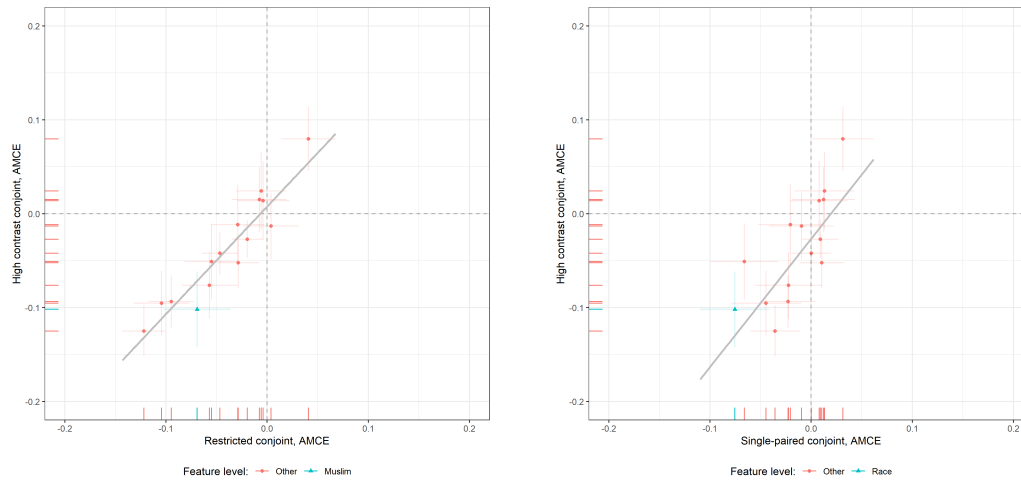
Building on the common assumption that respondents give more desirable answers when they anticipate that a sensitive feature is the main research objective, we would expect respondents to give different answers across conditions. More specifically, respondents should be more favorable to the black political candidates as well as the Muslim immigrants in the high-contrast design relative to the restricted paired design and the single-profile design.

Does design variation affect respondents' behaviour?

Before statistically testing the differences across designs, the AMCEs from the high-contrast designs are compared with the alternative designs in four scatterplots.¹⁰ A traditional visualization of the results from each study including attribute level-names are reported in Appendix C. Notice that the single-profile conjoint designs rely on a different type of task (evaluating one profile at a time instead of choosing between two) and a rating-based outcome measure. When analyzing the single-profile design, I use the ratings to code a binary variable as 1 if the rating is above the midpoint and 0 otherwise as is standard in the literature (Hainmueller and Hopkins 2015). The estimates in the single-profile designs are generally smaller, which implies that the unweighted effect estimates are not immediately comparable with the paired designs. Yet, the magnitude of the AMCE of the sensitive feature levels relative to the other AMCEs is directly comparable in the scatterplots.

Figure 3 depicts the results from Study 1. The left plot shows each coefficient estimate for the high-contrast design versus estimates obtained from the restricted design. The right plot shows each coefficient estimate from the high-contrast design versus estimates obtained from the single-profile design. Each point represents an AMCE-estimate with 95 percent confidence intervals with the coefficients ordered by their magnitude from most negative to most positive. Thus, the figure visualizes the extent to which larger AMCEs in the high-contrast designs are associated with larger effects in the alternative designs. In the same way, Figure 4 compares the estimates from study 2 when comparing the high-contrast design versus each estimate obtained from the restricted design (left side) and the single-profile design (right side). Altogether, there are no apparent differences in the AMCEs between the experimental conditions. Generally, the different designs yield highly comparable results. Importantly, this is also the case when comparing the AMCE-estimates of the sensitive features across designs (estimates for the sensitive feature levels are colored blue in the figures).

¹⁰ The analysis was conducted using R package version 0.3.1 (Leeper 2019).

Figure 3. Study 1. High-contrast estimates versus estimates from the alternative designs**Figure 4.** Study 2. High-contrast estimates versus estimates from the alternative designs

Note: The figures show each AMCE-estimate from the high-contrast design versus estimates obtained from the restricted (left) and the single-profile designs (right). Each point represents an AMCE-estimate with 95 percent confidence intervals with the coefficients ordered by their magnitude from most negative to most positive.

Next, I turn to a statistical comparison of the results. The outcome of interest is the differences in effects of the sensitive feature levels across designs. First, the two paired conjoint experiments in each study are compared. The paired designs rely on the same outcome and are therefore directly comparable. Hence, the effect of reducing attention to the the sensitive topic can be tested in a difference-in-difference model. In other words, I interact a design dummy variable (high-contrast = 0 / restricted = 1) with the sensitive topic in each study respectively.¹¹ A positive estimate indicates that respondents give more desirable answers in the high-contrast design which aligns with the expectation that SDB can be introduced by raising awareness to the sensitive feature. As shown in Figure 5 (a) the difference in the effects of the sensitive feature between designs is remarkably close to zero in both studies.¹² The effect of being black (Study 1) or Muslim (Study 2) is identical across designs.

In Figure 5 (b) I follow the same strategy in order to compare the high-contrast paired design and the single-profile design. However, since the experiments rely on outcomes measured on different scales, the comparison is not as straightforward. Since the AMCEs are consistently smaller in the single-profile designs, the size of the effect of a candidate being Black or Muslim is naturally smaller compared to the paired designs. I account for this by weighting the AMCEs in the single-profile design using the relative difference of all AMCEs between the single-profile and the paired designs as a weight.¹³ Again, there are no substantial differences between the single-profile and the high-contrast paired designs either as evidenced from Figure 5 (b).¹⁴

¹¹ For example, the estimand comparing the two paired design is expressed as:

$$(E[\text{choice} \mid \text{Black} \ \& \ \text{High-contrast}] - E[\text{choice} \mid \text{White} \ \& \ \text{High-contrast}]) \\ - (E[\text{choice} \mid \text{Black} \ \& \ \text{Restricted}] - E[\text{choice} \mid \text{White} \ \& \ \text{Restricted}])$$

An equivalent estimand is used in study 2 where "Black" equals "Muslim" and "White" equals the reference category.

¹² In study 1, the effect of being black compared to white increases the probability that a profile is chosen by 0.089 (SE = 0.016) in the high-contrast design and 0.092 (SE = 0.013) in the restricted paired conjoint. The effect of being Muslim is negative in both the high-contrast design with a coefficient of -0.096 (SE = 0.020) and -0.070 (SE = 0.017) in the restricted paired design.

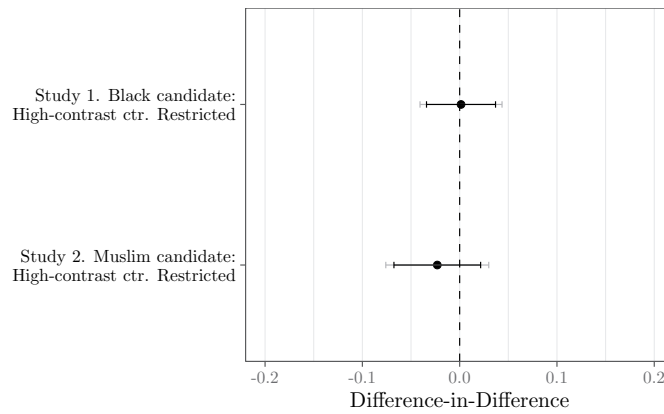
¹³ Specifically, the paired designs give AMCEs that are on average larger by a factor 1.87.

¹⁴ Note that the difference-in-differences are insignificant also without weighting the AMCEs from the Single-profile. See details in appendix F.

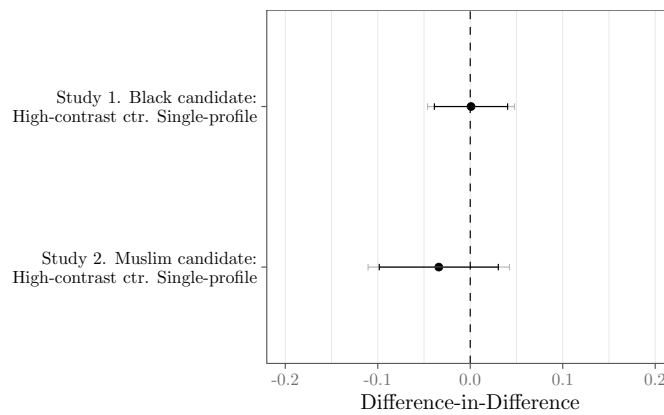
In summary, the results show that even when respondents anticipate a sensitive feature as important and at the same time have optimal conditions for tailoring their answers, it does not change their responses.

Figure 5. Difference-in-differences

(a) High-contrast paired design vs. Restricted paired design



(b) High-contrast paired design vs. Single-profile design



Note: The figures show differences in effects of the sensitive levels between (a) the high-profile paired design and the restricted paired design and (b) the high-profile paired design and the single-profile design.

What else could explain the null-findings?

The results are supposedly good news to researchers conducting conjoint experiments: we should not be too concerned with implementing designs that, at least in theory, increase the risk of SDB. In this section I test alternative explanations for the null-findings.

First and foremost, one concern is that (some) respondents would disagree that it is socially desirable to have a preference against black or Muslim candidates.¹⁵ Another methodological objection is that the increased awareness to a sensitive feature also introduces demand-effects that cancels out SDB. Demand effects are caused by respondents attempting to validate a researcher's hypothesis by behaving in line with what they perceive as the expected behaviour (Mummolo and Peterson 2018). If respondents anticipated that the present study expected to find a bias against black or Muslim profiles, they might have answered in a way that would "help" the researcher confirm the hypothesis, which would bias the effect in the opposite direction than SDB.

To bolster the results, I therefore rerun the analysis in subsets of the samples where social desirability pressures related to preferences regarding Race and Muslim affiliation are arguably more pronounced. Firstly, political liberals have been found to be more likely to give untruthful answers to questions regarding race (Gilens, Sniderman, and Kuklinski 1998) and immigration restrictionist policy questions (Janus 2010). In both studies, I reestimate the difference-in-differences between the paired designs in subsets of respondents that identify as liberal on the pre-treatment questions.¹⁶ As demonstrated in the supplementary material (appendix D), the difference-in-differences from the liberal subset is a precisely estimated zero (Study 1: -0.014, SE = 0.031; Study 2: 0.0007, SE = 0.049). Secondly, to further bolster the results, I look at subsets of the samples that are more likely to be attentive to self-presentational concerns – and thus where we would expect SDB to be most pronounced. Previous studies have found that high self-monitors are more likely to give appropriate answers to sensitive questions. Following Berinsky and Lavine (2011), study 2 included three items from the self-monitoring scale that was also used in a conjoint analysis by Hainmueller and Hopkins

¹⁵ Although M-turkers tend to be younger and more liberal compared to a nationally representative samples (Levay, Freese, and Druckman 2016).

¹⁶ Respondents with a score >6 on a 0-10 scale ranging from "Very conservative" to "Very liberal".

(2015).¹⁷ As shown in the supplementary material (appendix D), the difference-in-differences from the high-monitor subset is close to null in study 2 (.020, SE = 0.043).

A third concern is, that the "treatment" in the high-contrast paired designs was not assigned before the outcomes were measured, but rather is embedded in the design. Hence, it is possible that the sensitive dimension became increasingly obvious to respondents as they worked their way through the five conjoint pairs. In other words, respondents assigned to the high-contrast paired design could have been more aware about the sensitive feature when they were asked to choose between a black and a white candidate for the third, fourth and fifth time. In that case, results should change towards more politically correct answers towards the end of the experiment. To test this, I compare estimates in the high-contrast designs from pair 1-5 respectively. The changes in effect sizes as respondents answer the five pairs in the high-contrast designs are inconsequential and does not support the notion that respondents change preferences as the contrast on a sensitive dimension is repeated (See appendix E in the supplementary material).

Finally, we might worry that the semi-professional respondents on Amazon's Mechanical Turk are somehow less prone to social desirability pressures than population based samples. The experiments provided in this study cannot shed light on this concern. However, research indicates that survey experiments conducted on convenience samples like M-turkers yield similar effects as those from national probability samples (Berinsky, Huber, and Lenz 2012; Coppock 2018; Mullinix et al. 2015). Furthermore, since a lot of social science conjoint experiments are carried out in convenience respondent pools such as M-turk, examining the research question in a convenience sample has a value in itself.

¹⁷ The following questions are used: "When you're with other people, how often do you put on a show to impress or entertain them?" Response categories: Always, Most of the time, About half the time, Once in a while, Never. "How good or bad of an actor would you be?" Response categories: 'Excellent', 'Good', 'Fair', 'Poor', 'Very poor'. "When you are in a group of people, how often are you the center of attention?" Response categories: 'Always', 'Most of the time', 'About half the time', 'Once in a while', 'Never'.

Conclusion and discussion

Conjoint designs are often claimed to limit concerns over social desirability bias: that research subjects are biased towards normatively positive responses. This is based on two arguments: due to the large number of features, (i) respondents cannot infer the main intent of the experiment, and (ii) they can easily justify inappropriate answers. However, the extent to which these arguments hold depends on the specific type of conjoint experiment employed. The present study tests the importance of design variations by comparing answers across different types of conjoint designs.

The results provide evidence that the design of conjoint experiments have an effect on respondents' inferences about the main objective: respondents pay significantly more attention to sensitive features in a paired conjoint design where the sensitive feature levels are frequently contrasted compared to designs where the contrast is less obvious (single-profile and paired conjoints with restricted randomization schemes). However, the core quantities of interest are remarkably stable across designs, suggesting that the substantive conclusions are not threatened by the specific choice of design.

There are several implications of these results. First, while this study cannot rule out that SDB is ever an issue in conjoint experiments, it is reassuring that different types of conjoint designs give the same results. Second, the stability of the results across designs also goes against recent suggestions that paired-conjoint designs makes it easier for respondents to act 'strategically' in order to provide desirable answers (Mummolo and Peterson 2018). There is no evidence that respondents act differently when presented with a within-subject design compared to a between-subject study. Thirdly, and consequently, there is no immediate reason to choose a design that is otherwise sub-optimal in order to disguise sensitive topics. Had this study proved that respondents' priorities change when respondents recognize sensitive features as important, the implications would be serious. As a main concern, it would question the inferences that researchers are able to make from conjoint designs more generally. Moreover, it would emphasize the need to choose otherwise sub-optimal designs in order to downplay sensitive features. The results presented in this article indicates that there is no reason that researchers using conjoint designs should limit the number

of pairs or restrict the probabilities of certain feature levels and thereby decrease statistical power and/or hamper external validity. Nor is there strong arguments for using single-profile designs unless they are preferable for other reasons. Finally, beyond conjoint designs specifically, the present study supports recent evidence by Mummolo and Peterson (2018) and De Quidt, Haushofer, and Roth (2017) that researchers should not be too concerned with respondents' awareness to research intentions in survey experiments.

References

- Arceneaux, Kevin (2008). “Can partisan cues diminish democratic accountability?” In: *Political Behavior* 30.2, pp. 139–160.
- Bansak, Kirk, Jens Hainmueller, and Dominik Hangartner (2016). “How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers”. In: *Science*, aag2147.
- Bansak, Kirk, Jens Hainmueller, Daniel J Hopkins, et al. (2017). “Beyond the breaking point? Survey satisficing in conjoint experiments”. In: *Political Science Research and Methods*, pp. 1–19.
- Berinsky, Adam J, Gregory A Huber, and Gabriel S Lenz (2012). “Evaluating online labor markets for experimental research: Amazon. com’s Mechanical Turk”. In: *Political analysis* 20.3, pp. 351–368.
- Berinsky, Adam J and Howard Lavine (2012). “Self-monitoring and political attitudes”. In: *Improving public opinion surveys: Interdisciplinary innovation and the American national election studies*, pp. 27–45.
- Broockman, David and Evan Soltas (2017). *A natural experiment on taste-based racial and ethnic discrimination in elections*.
- Bullock, John G (2011). “Elite influence on public opinion in an informed electorate”. In: *American Political Science Review* 105.3, pp. 496–515.
- Carnes, Nicholas and Noam Lupu (2016). “Do voters dislike working-class candidates? Voter biases and the descriptive underrepresentation of the working class”. In: *American Political Science Review* 110.4, pp. 832–844.
- Coppock, Alexander (2018). “Generalizing from survey experiments conducted on Mechanical Turk: A replication approach”. In: *Political Science Research and Methods*, pp. 1–16.
- De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth (2017). *Measuring and Bounding Experimenter Demand*. Tech. rep. National Bureau of Economic Research.
- Dickson, Eric (2011). “Economics vs. Psychology Experiments”. In: *The Handbook of Experimental Political Science*. Cambridge University Press.

- Gilens, Martin, Paul M Sniderman, and James H Kuklinski (1998). “Affirmative action and the politics of realignment”. In: *British Journal of Political Science* 28.1, pp. 159–183.
- Groves, Robert M et al. (2011). *Survey methodology*. Vol. 561. John Wiley & Sons.
- Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto (2015). “Validating vignette and conjoint survey experiments against real-world behavior”. In: *Proceedings of the National Academy of Sciences* 112.8, pp. 2395–2400.
- Hainmueller, Jens and Daniel J Hopkins (2015). “The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants”. In: *American Journal of Political Science* 59.3, pp. 529–548.
- Hainmueller, Jens, Daniel J Hopkins, and Teppei Yamamoto (2014). “Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments”. In: *Political Analysis* 22.1, pp. 1–30.
- Hariri, Jacob Gerner and David Dreyer Lassen (2017). “Income and outcomesSocial desirability bias distorts measurements of the relationship between income and political behavior”. In: *Public Opinion Quarterly* 81.2, pp. 564–576.
- Holtgraves, Thomas (2004). “Social desirability and self-reports: Testing models of socially desirable responding”. In: *Personality and Social Psychology Bulletin* 30.2, pp. 161–172.
- Horiuchi, Yusaku, Daniel M Smith, and Teppei Yamamoto (2017). “Identifying Voter Preferences for Politicians’ Personal Attributes: A Conjoint Experiment in Japan”. In:
- Janus, Alexander L (2010). “The influence of social desirability pressures on expressed immigration attitudes”. In: *Social Science Quarterly* 91.4, pp. 928–946.
- Kam, Cindy D (2007). “Implicit attitudes, explicit choices: When subliminal priming predicts candidate preference”. In: *Political Behavior* 29.3, pp. 343–367.
- Kaminska, Olena and Tom Foulsham (2013). *Understanding sources of social desirability bias in different modes: Evidence from eye-tracking*. Tech. rep. ISER Working Paper Series.
- Kirkland, Patricia A and Alexander Coppock (2017). “Candidate Choice Without Party Labels”. In: *Political Behavior*, pp. 1–21.

- Krupnikov, Yanna and Adam Seth Levine (2014). “Cross-sample comparisons and external validity”. In: *Journal of Experimental Political Science* 1.1, pp. 59–80.
- Krupnikov, Yanna, Spencer Piston, and Nichole M Bauer (2016). “Saving Face: Identifying Voter Responses to Black Candidates and Female Candidates”. In: *Political Psychology* 37.2, pp. 253–273.
- Krysan, Maria and Mick P Couper (2003). “Race in the live and the virtual interview: Racial deference, social desirability, and activation effects in attitude surveys”. In: *Social psychology quarterly*, pp. 364–383.
- Kuklinski, James H et al. (1997). “Racial prejudice and attitudes toward affirmative action”. In: *American Journal of Political Science*, pp. 402–419.
- Leeper, Thomas J. (2019). *cregg: Simple Conjoint Analyses and Visualization*. R package version 0.3.1.
- Levay, Kevin E, Jeremy Freese, and James N Druckman (2016). “The demographic and political composition of Mechanical Turk samples”. In: *Sage Open* 6.1, p. 2158244016636433.
- Lewis-Beck, Michael S, Charles Tien, and Richard Nadeau (2010). “Obama’s missed landslide: a racial cost?” In: *PS: Political Science & Politics* 43.1, pp. 69–76.
- Liu, Hanzhang (2018). “The Logic of Authoritarian Political Selection: Evidence from a Conjoint Experiment in China”. In: *Political Science Research and Methods*, pp. 1–18.
- McDermott, Rose (2002). “Experimental methods in political science”. In: *Annual Review of Political Science* 5.1, pp. 31–61.
- Mullinix, Kevin J et al. (2015). “The generalizability of survey experiments”. In: *Journal of Experimental Political Science* 2.2, pp. 109–138.
- Mummolo, Jonathan and Erik Peterson (2018). “Demand effects in survey experiments: An empirical assessment”. In: *American Political Science Review*, pp. 1–13.
- Ono, Yoshikuni and Masahiro Yamada (2016). “Do Voters Prefer Gender Stereotypic Candidates?: Evidence from a Conjoint Survey Experiment in Japan”. In:
- Sen, Maya (2017). “How political signals affect public support for judicial nominations: Evidence from a conjoint experiment”. In: *Political Research Quarterly* 70.2, pp. 374–393.

- Teele, Dawn Langan, Joshua Kalla, and Frances Rosenbluth (2018). “The Ties that Double Bind: Social Roles and Women’s Underrepresentation in Politics”. In: *American Political Science Review*, pp. 1–17.
- Tourangeau, Roger and Ting Yan (2007). “Sensitive questions in surveys.” In: *Psychological bulletin* 133.5, p. 859.

Supplementary material: Social desirability bias in conjoint experiments:
What is the optimal design when studying sensitive topics?

Appendix A. Constructing and fielding the conjoint experiments

The experiments were implemented in Qualtrics and fielded at Amazon’s Mechanical Turk. The sampling took place between August 8 and August 30. I used the build-in randomize option in Qualtrics. Only respondents who answered all questions are included in the final sample. The respondents were presented with a paired design or a single-profile design. Screenshots of a paired conjoint design and a single-profile conjoint design are shown in Figure A1 and A2.

A1. Example of discrete choice conjoint

Candidate A		Candidate B
66	Age	46
Male	Gender	Female
Seek better job in U.S.	Reason for application	Escape political/religious persecution
Other	Religion	Muslim
Nurse	Profession	Nurse
5+ years	Job experience	1-2 years
Entered the U.S. once before on a tourist visa	Prior trips to the US	Entered the U.S. once before without legal authorization

Please indicate which of the two immigrants you would personally prefer to see admitted to the United States

☐ Candidate A

☐ Candidate B

A2. Example of rating-based conjoint

	Candidate
Age	53
Gender	Male
Race	White
Religion	Jewish
Education	Albany Law School
Previous work experience	Public defender
Clerkship experience	Served as law clerk

Where would you place your level of support for this potential candidate?

1. Strongly oppose	2	3	4. Neither	5	6	7. Strongly support
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix B. Features and feature levels

Table B give details on the features and feature levels used to generate the profiles in Study 1 and Study 2.

Table B. Features and feature levels included in the conjoint experiments

Study 1		Study 2	
Feature	Level	Feature	Level
Gender	Male, Female	Gender	Male, Female
Age	25-75 (Continuous)	Age	25-75 (Continuous)
Race	White, Black	Religion	Atheist, Protestant, Jewish, Muslim, Catholic, Other
Education	Yale Law School, Florida State Uni., Albany Law School	Reason for application	Seek better job, Reunite with family members, Escape religious/political persecution
Religion	Mormon, Mainline Protestant, Jewish, Evangelical Protestant, Catholic	Profession	Doctor, Nurse, Teacher, Waiter, Construction worker, Computer programmer
Clerkship experience	Did not serve as law clerk, did serve as law clerk	Working experience	None, 1-2 years, 3-5 years, More than 5 years
Previous work experience	elected politician, law professor, lawyer in private practice, non-profit lawyer, public defender	Prior trips to the U.S.	Never been to the U.S., Spent six months with family, Visited once without legal authorization, Visited once on tourist visa, Visited many times on tourist visa

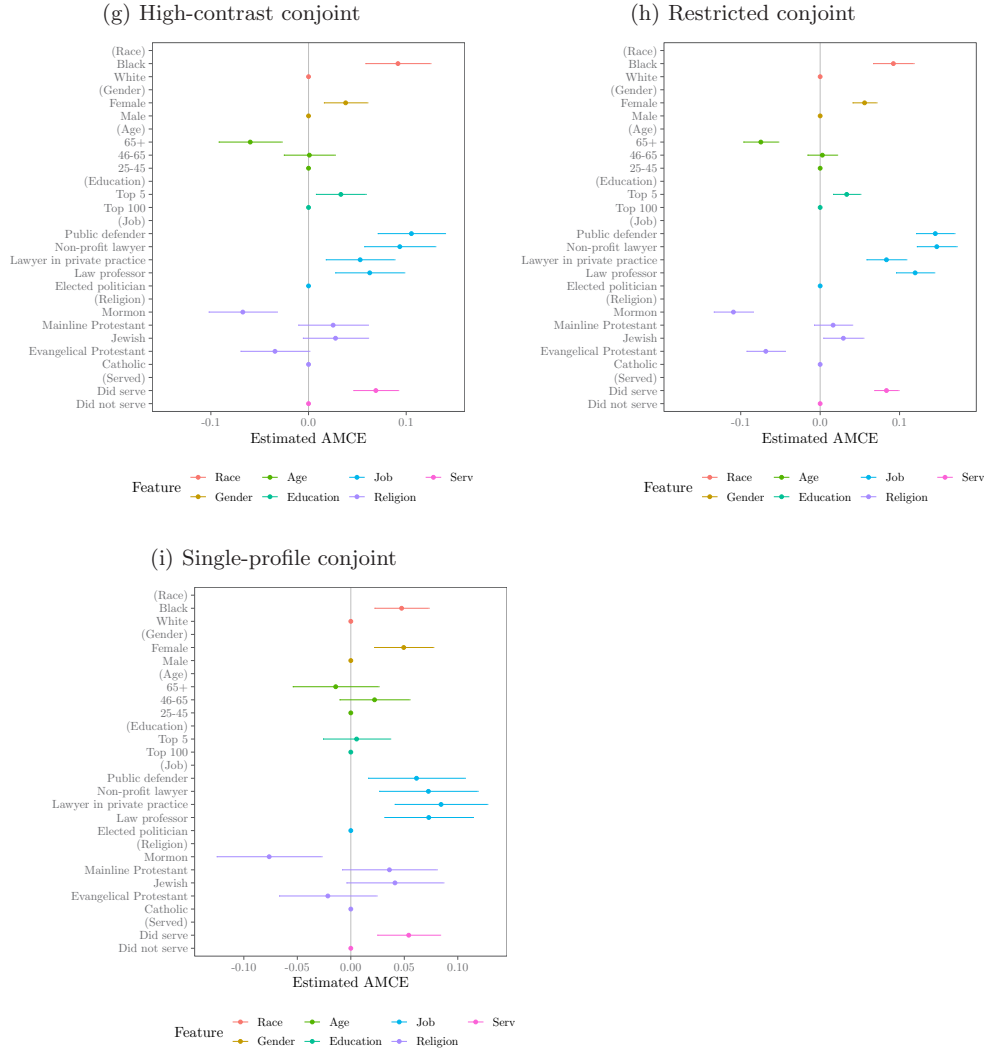
Appendix C. Results from conjoint experiments

The AMCEs from each study are visualized in Figure C.¹⁸ Notice that the single-profile conjoint designs rely on a different type of task (evaluating one profile at a time instead of choosing between two) and a rating-based outcome measure. When analyzing the single-profile design, I use the ratings to code a binary variable as 1 if the rating is above the midpoint and 0 otherwise as is standard in the literature (Hainmueller and Hopkins 2015). This implies that the unweighted effect estimates are not immediately comparable with the paired designs. Yet, the magnitude of the AMCE of the sensitive feature levels relative to the other AMCEs are strikingly similar across the three conditions in both studies. In study 1, the effect of being black is positive and significant in all of the three conditions.¹⁹ In study 2, the effect of being Muslim is negative and significant in all conditions. Altogether, there are no apparent differences in the core quantities of interest between the three experimental conditions.

¹⁸ The AMCE represents the marginal effect of a given attribute averaged over the joint distribution of the remaining attributes. Standard errors are corrected for within respondent clustering.

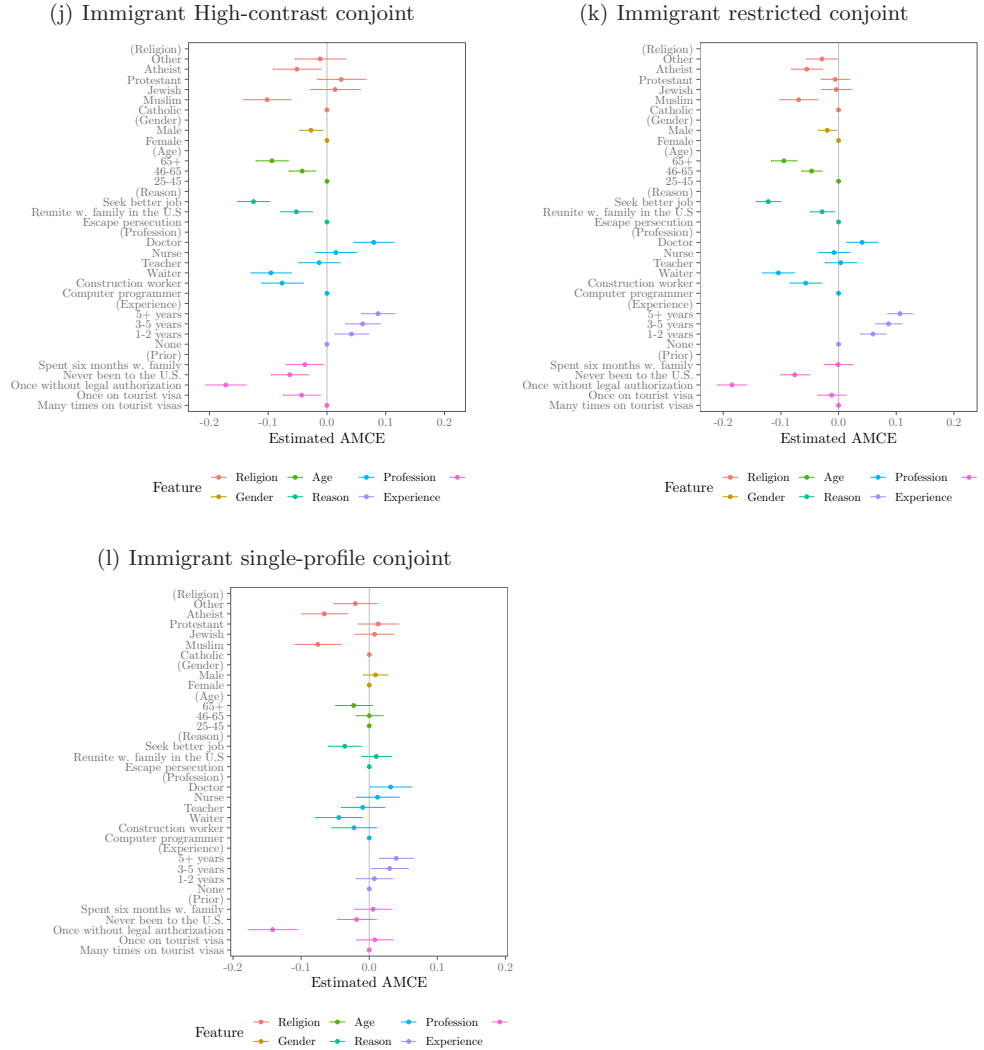
¹⁹ This is unsurprising considering evidence from previous candidate choice experiments and the fact that this experiment did not include political leaning (which can crowd out effects of demographic characteristics).

Figure C (i). Results from Supreme Court candidate conjoint experiments (N=3,493)



Note: Each estimate represents the effect of a given feature level compared to a reference level when averaging over the joint distribution of the remaining features.

Figure C (ii). Results from immigrant conjoint experiments (N=3,566)



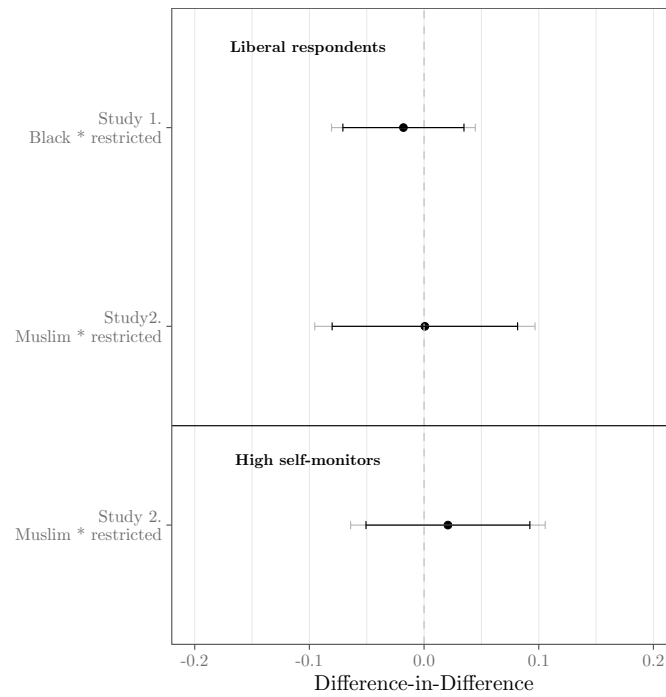
Note: Each estimate represents the effect of a given feature level compared to a reference level when averaging over the joint distribution of the remaining features.

Appendix D. Robust to demand effects

One way to further bolster the results is to look at a subset of the sample that are more attentive to self-presentational concerns. First, I reestimate the difference-in-differences between the two paired designs in both studies comparing respondents across score on a 0-10 scale ranging from "Very conservative" to "Very liberal". Political liberals have been found to be more likely to give untruthful answers to questions regarding race (Gilens, Sniderman, and Kuklinski 1998) and immigration restrictionist policy questions (Janus 2010), and we would therefore expect to see stronger SDB among liberals. Secondly, previous studies have found that high self-monitors are more likely to give appropriate answers to sensitive questions. Following Berinsky and Lavine (2011), I used three items from the self-monitoring scale that was also used by Hainmueller and Hopkins (2015).²⁰

²⁰ The following questions are used: "When you're with other people, how often do you put on a show to impress or entertain them?" Response categories: Always, Most of the time, About half the time, Once in a while, Never. "How good or bad of an actor would you be?" Response categories: Excellent, Good, Fair, Poor, Very poor. "When you are in a group of people, how often are you the center of attention?" Response categories: Always, Most of the time, About half the time, Once in a while, Never.

Figure D. Difference-in-differences among subsets of respondents

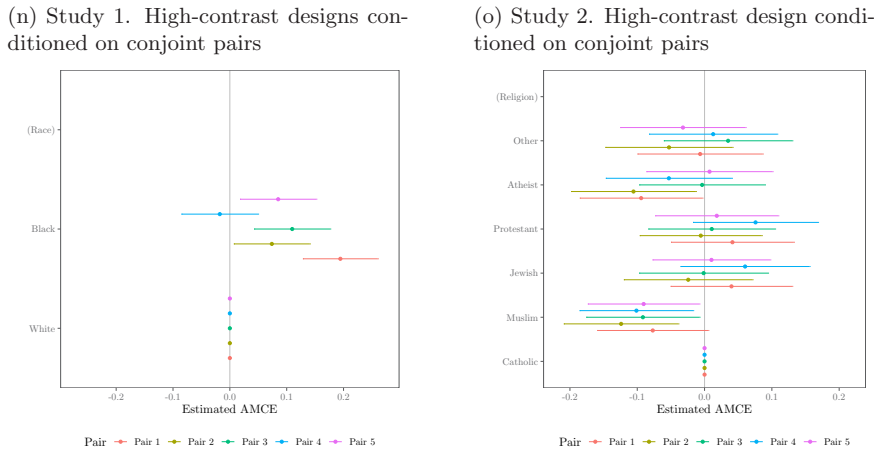


Note: Difference-in-differences between the paired designs when including only liberal respondents (Study 1 and Study 2) and when including only high self-monitors.

Appendix E. AMCEs across repeated choices

The treatment was not assigned before the experiments, but rather is embedded in the design. Hence, respondents in the paired design with high contrast may have been more aware about the sensitive feature when they were asked to choose between a black and a white candidate for the third, fourth and fifth time. To test this, I compare estimates in the high-contrast designs from pair 1, 2, 3, 4 and 5 respectively. As shown in Figure C, the change in effect sizes as respondents answer the five pairs in the high-contrast designs are inconsequential and the results do not support the notion that respondents change preferences as the contrast on a sensitive dimension is repeated.

Figure E. Effect estimates across the five conjoint pairs



Appendix F. Difference-in-difference with and without weighting

The single-profile design and the paired designs are not directly comparable. Firstly, the tasks that respondents were asked to solve differ: in the paired designs they are presented with two profiles while in the single-profile, they only see one at a time. Moreover, the outcome measure is different as well (either a forced choice or rating). Secondly, the AMCEs in the single-profile conjoint designs are generally smaller compared to the paired designs. This makes the direct comparison of the effect

estimate of the sensitive feature across design problematic. One way to solve this is to re-weight the AMCEs of the single profile designs. In other words, I estimate the average difference in effect estimates of all *other* features between the single-profile and the paired designs. On that basis, the estimate of the sensitive feature is re-weighted. Across both studies, all other features than the sensitive are on average larger by a factor 1.87 relative to the effects in the single-profile conjoint. The paper gives the weighted difference-in-difference between the paired design with high contrast and the single-profile design. Table E gives the difference-in-difference both with and without weighting.

Figure F (i). Comparison of the high contrast paired design and the single profile designs with and without weighting

Study 2. AMCE of a candidate being black	
	Single-profile is weighted / unweighted
Paired, high contrast	0.089 (0.017)
Single-profile	0.088 (0.024) / 0.047 (0.012)
Difference-in difference	0.0008 (0.029) / 0.041 (0.022)

Figure F (ii). Comparison of the high contrast paired design and the single profile designs with and without weighting

Study 2. AMCE of a candidate being Muslim	
	Single-profile is weighted / unweighted
Paired, high contrast	-0.096 (0.020)
Single-profile	-0.131 (0.033) / -0.070 (0.017)
Difference-in difference	-0.034 (0.039) / 0.026 (0.027)