

Accurate, Efficient, and Immoral?
When Statistics Motivate Discrimination

Ida Bruun Nørregaard

Accurate, Efficient, and Immoral?
When Statistics Motivate Discrimination

PhD Dissertation

Politica

© Forlaget Politica and the author 2025

ISBN: 978-87-7335-335-6

Cover: Svend Siune

Print: Fællestrykkeriet, Aarhus University

Layout: Annette Bruun Andersen

Submitted September 27, 2024

The public defense takes place January 31, 2025

Published January 2025

Forlaget Politica

c/o Department of Political Science

Aarhus BSS, Aarhus University

Bartholins Allé 7

DK-8000 Aarhus C

Denmark

Table of Contents

Acknowledgments	7
Preface	11
Chapter 1: Introduction.....	13
1.1 Three Guiding Research Questions	14
1.2 The Central Argument.....	15
1.3 Recruitment as a Case	16
1.4 Roadmap	17
Chapter 2: Methodologies.....	19
2.1 Experimental (and Empirical) Political Science.....	19
2.2 Analytical Philosophy	20
2.3 The Fruitfulness Approach	22
Chapter 3: Theories of Discrimination	25
3.1 A Typology of Discrimination	26
3.2 The Moral Relevance of Intention, Statistics, and Accuracy	26
Chapter 4: When and Why do People Accept Recruitment Discrimination?	29
4.1 Methods and Research Design.....	29
4.2 Results.....	31
4.3 Reflections on the Research Question.....	37
Chapter 5: When and Why is Statistical Discrimination Morally Wrong?	39
5.1. Prominent Objections to Statistical Discrimination.....	39
5.2 Unchangeable by Will	41
5.3 Criteria for Wrongful Objectification	42
5.4 The Framework's Scope.....	43
5.5 Reflections on the Research Question and Implications	44
Chapter 6: Reducing the Acceptance of Statistical Discrimination	47
6.1 Informational and Moral Frames	48
6.2 Method and Research Design.....	49
6.3. Results.....	51
6.4 Reflections on Moral Frames and the Research Question.....	53
Chapter 7: Discussion and Conclusion	55
7.1 Normativity, Empirics, and Paternalism	56
7.2 Limitations and Future Research	58
7.3 Social and Political Implications	61
English Summary	63
Dansk Resumé.....	65
Bibliography	67

Acknowledgments

It takes a village. Not only to raise a kid (which people much smarter than me tend to say) but also to raise a PhD. This final result, the dissertation, is not only the fruit of my labor. It is the outcome of three years of eminent supervisors, inspiring colleagues, supportive friends, and a loving family. I want to spend some time acknowledging and, most importantly, saying thank you to the village that helped me grow on my PhD journey.

To my supervisors, I want to express how grateful I am for your support and guidance throughout. It goes without saying that you both have contributed with excellent theoretical and methodological insights. But maybe more importantly, you have encouraged me to follow my interests, ideas, and ambitions. Thank you, Andreas, for providing an optimistic, calm, and nuanced perspective whenever academic life (whether that be reviewer comments, career, or other challenges) seemed a little chaotic. And thank you, Martin, for always approaching my ideas with energy and enthusiasm and knowing when to push me a little extra. I could not have wished for a more balanced supervisor team. And for future reference, I strongly recommend 10/10!

I was not the only academic kid raised in this village. I had the pleasure of sharing an office with one such kid who is truly one of a kind. I have shared my office with you, Jannik (one could call you my academic twin), from start to finish. From day one, I had someone to share the PhD experience with. Not only do I feel extremely lucky that I was paired with someone empathetic, cheery, and funny, but I also feel privileged by your tolerance of a (sometimes too) talkative office mate. I cannot count the times I have said, “Okay, one last thing, and then I promise I will let you work.” Yet every single time, my interruption was met with interest and enthusiasm about what I had to say. You have been a great colleague and friend both in times of celebration and in times of challenges. Thank you!

My village also consisted of a bunch of brilliant people from the political theory section and CEPDISC. In an academic world where competition and uncertainty are part of “the game,” the political theory section offers a refuge. Thank you for being genuinely kind people. I believe and hope that those who shape you academically leave a lasting footprint. You are all great examples of how to be generous, supportive, and warm (and without a doubt competent) colleagues. If that is not a competitive advantage, I do not know what is. An extra thanks to Søren for our many insightful chats, whether that is about discrimination or absurd dilemmas; to Viki for all the times I popped in unannounced with questions and for your support in looking for a future in academia; to Marion for being enthusiastic about my research (it is infectious);

and to Simone, Ida V, Lydia, Jake, Rasmus H, Andreas B, Miklós, Anna, and Jens for being great colleagues and friends whom I am happy to have gotten to know and spent time with—whether that be in section meetings, in book clubs, or on the dance floor!

I am also very happy to have been part of CEPDISC, which has enabled me (and us collectively) to nerd out on an extremely important and yet complicated topic (discrimination). The center and its people are a testament to how much we can learn from each other across research fields and methods. Thank you, Kasper, for the opportunities you create in CEPDISC (and for supervising my master's thesis which sparked my interest in experimental philosophy). A huge thank you to Maj—you have been a lifesaver countless times and make the CEPDISC world go round! To the current and former CEPDISC PhDs, Simone, Astrid, MJ (I will miss our yapping sessions), Benedicte, Alexandra, Lydia (we must catch up on Greek dancing), and Samuel, are such an extraordinary bunch, and I cannot wait to see the great places you will go!

Whereas any journey has its challenges, I believe my biggest challenge is yet to come. This will be realizing that ending the PhD journey means that I will no longer have the same opportunity to spend every lunch and coffee break with the phenomenal people who make up (and used to make up) the PhD group. I was lucky to start my journey surrounded by fantastic people like Jannik, Lea, Ane, Mathias J, Anna V, Ida A, Daniel, William, Nadine, Karl, Rikke, Ashraf, Simone, Nanna, Jens, and Aske. And equally lucky with the people I got to know on the way, such as Louise, Ida V, Nic, MJ, Jenny, Maria, Benedicte, Rasmus H, and Mario. I will miss the unique understanding and solidarity of the PhD group, the fun parties, cozy board game afternoons, forest walks, and the best book club in town (a shoutout to my fellow book nerds: *The Idas*, Aske, Ashraf, Maria, Signe, Ali, MJ, Oksana, and Anna). And to the spunkiest, most mysterious, demure, podcast-loving, finger-dancing bunch I have ever met—Hannah, Tim, Rasmus (Razzy G), Mathies, Mathias, Leon, Edoardo—I think there is only one way I can sum up how I feel: How's Italy?

Even a village with a strong spirit needs some formal structures to make life run smoothly. During my PhD, three excellent PhD coordinators have been part of this structure, making sure that the journey has been as smooth as possible. Thank you, Christoffer, Helene, and Lasse, for being a solid backbone in the PhD group. Thank you to the amazing TAP at the department, who keep cool, calm, and collected when we academics have our heads in the sky. A special thank you to Njall, who always makes everyone feel seen.

While the PhD journey has brought amazing experiences, connections, and friends, it has also been important for me to step away from academic life from time to time. Outside of academic life and work, I have an outstanding group of friends. I love spending time with you, learning from you, laughing

with you, and growing with you. Camilla, Sigrid, Pernille, Thea, Anna, Ro, and Lasse—thank you for being a breath of fresh air!

To those who raised me long before my PhD journey began, my family, I am grateful for your unconditional love and support. I feel lucky to have parents and bonus parents who have encouraged me to pursue the paths in life I was curious about (and sometimes a bit scared of). Also, a special thank you to my grandparents, who have inspired me tremendously. Thank you for instilling a passion for learning about and discovering the world and showing me how to treat people with kindness and patience. To one of the best people I know, my sister Eva—during my PhD, we have shared a home. A home of laughter, dreams, and absolute silliness. Whether I am living in the room next to you or we are 1,000 kilometers apart, my home will always be where you are! And to you, Matthias, I owe much of my sanity. Without your selflessness, care packages, and spontaneous adventures, my PhD journey would have looked very different. And all of this while you are doing a PhD of your own. I believe I have stumbled upon a real superhero.

Finally, I want to end by saying thank you to my cat, Momo. Thank you for your indifference to titles and accomplishments, for reminding me—in that wonderfully ruthless way only cats can—that what truly matters is not what we achieve, but how we treat those around us.

Thank you all for being my village!

Ida

Aarhus, January 2025

Preface

My PhD dissertation, titled “Accurate, Efficient, and Immoral? When Statistics Motivate Discrimination,” consists of a summary report and the three articles listed below. The summary report introduces the dissertation’s core research question, methodological approaches, and theoretical framework, and it summarizes the main results and arguments based on the articles. Details of the theories, methods, and analyses are provided in the individual articles, which I refer to as Articles A, B, and C.

Article A: “When Do People Accept Discrimination? An Experimental Assessment of the Persuasiveness of Statistics.” R&R at *Political Behavior*.

Article B: “You are Nothing but a Number: The Objectification Problem of Statistical Discrimination.” *Under review*.

Article C: “The Moral Foundations of Discrimination: How Conservative Values Persuade and Liberal Values May Polarize.” *Under review*.

Chapter 1: Introduction

Discrimination has long been a salient and politicized societal issue. Despite the development and expansion of anti-discrimination legal frameworks and social norms, discriminatory practices and attitudes prevail in many democratic and egalitarian societies (Auspurg et al., 2019; Beatty & Sommervoll, 2008; Stefan et al., 2018; Zschirnt & Ruedin, 2016). This has puzzled politicians, citizens, and researchers alike. Why does discrimination still constitute a widespread social phenomenon when most institutions and citizens seem to condemn it?

Researchers have identified psychological antecedents, historical and institutional structures, and social group dynamics as possible explanations of contemporary discrimination (Al Ramiah et al., 2010; Fibbi et al., 2021; Roth et al., 2019). In this dissertation, I point to an alternative explanation—one centering on people’s moral perceptions of discriminatory practices and how different types of discrimination cause different moral values to be salient in people’s normative judgments, including value conflicts or trade-offs. One type of discrimination that exemplifies the perceived moral dilemmas or trade-offs in discriminatory practices is what is termed *statistical discrimination*. Unlike what many people might associate with discrimination, statistically discriminatory practices do not stem from animosity or distaste for certain groups. Rather, they are motivated by efficiently pursuing a presumed legitimate goal (Schauer, 2018). To pursue that goal, the discriminator relies on statistical group differences as indicators of which group of people should be allocated resources, excluded, receive benefits, and so forth. An example of statistical discrimination could be when employers reject female applicants because they are more likely to call in sick than men, or when employers reject Muslim applicants because they believe they are less productive during Ramadan. In these and similar cases of discrimination, the statistical group differences make efficiency and workplace loyalty salient values in people’s minds that need to be weighed against the value of, say, equal treatment.

In my dissertation, I examine the moral values at stake in discriminatory practices with an emphasis on statistical group differences as a trigger of moral value conflict. In other words, I examine whether the existence of statistical group differences between the parties affected by discrimination triggers a perception of value conflict or value trade-offs that is not present when statistical group differences are absent. The overarching research question that I answer is thus: *What role do statistical group differences play in the*

moral acceptability of discrimination? My focus lies primarily within the domain of recruitment discrimination, as this arena faces a great deal of political pressure and attention. Currently, established recruitment practices are being scrutinized and reevaluated, for instance in light of affirmative action initiatives where debates about discrimination are at the forefront. Recruitment practices in both private and public institutions that consider statistical group differences and disparities have become politically contentious and polarizing (Bell, 2021; DeBell, 2017). Recruitment is thus a highly relevant case for examining the role of statistical group differences and moral value conflicts. I will elaborate on why recruitment discrimination is a fitting case for investigating the role of statistical group differences in Section 1.4.

1.1 Three Guiding Research Questions

To answer the overarching research question, the dissertation is structured by three guiding questions. These three questions cover different theoretical and methodological aspects of the dissertation and enable a nuanced and interdisciplinary exploration of the overarching research question.

An important first step in answering the dissertation's core question is to test what makes people accept discrimination in the first place. This motivates the first guiding question of the dissertation: *When and why do people accept recruitment discrimination?* (Q1) This question investigates (i) whether people accept statistical discrimination to a greater extent than taste-based discrimination and if so, (ii) which feature of statistical discrimination drives people's acceptance. It also investigates (iii) the justifications people provide for accepting discrimination. To answer the first guiding question, I rely on the findings from Article A, in which I employed experimental methods to identify and analyze the causal relationship between the features of statistical discrimination and people's acceptance of discrimination. I also rely on the findings from Article C, in which I used quantitative and automated text analysis methods to identify the moral values people use to justify their moral assessments of discrimination.

If people tend to accept statistical discrimination more than, say, taste-based discrimination, it prompts the question of whether we should then conclude that statistical discrimination *is in fact* less morally worrisome than other types of discrimination. After all, the unequal and potentially unfair treatment involved in discrimination still looms large in the background. Regardless of statistical group differences, those targeted by a discriminatory action are disadvantaged because of the group to which they are perceived to belong and *not* because of their own character and behavior. As such, it is important to morally scrutinize discrimination based on accurate statistics. This

motivates the second research question of the dissertation, asking *what makes accurate statistical discrimination morally wrong when it is?* (Q2) Unlike the first guiding research question, this question is normative. It does not seek to uncover how people assess statistical discrimination and what their reasons are, but instead seeks to systematically identify what is morally at stake in cases of statistical discrimination. I rely on Article B to answer the second guiding question. In Article B, I use an analytical philosophical approach, analyzing and discussing the (potential) moral wrongdoings involved in statistically discriminatory cases.

Based on the insights into how the existence of statistical group differences affects people's acceptance of discrimination and the moral wrongdoing involved in most cases of statistical discrimination, the final guiding question concerns whether we can do anything to decrease people's acceptance of statistical discrimination. As the appeal to statistical group differences seems to prompt people to give more weight to loyalty and the perceived objectivity and efficiency of drawing on group distinctions rather than to the unequal and unfair treatment involved, one strategy could be to reemphasize the moral costs involved in statistical discrimination. These considerations motivate the third and final guiding research question: *How does morally reframing statistical discrimination influence people's acceptance hereof?* (Q3) To answer the third guiding question, I build on the experimental findings of Article C.

By addressing the three guiding research questions, this dissertation provides an answer to what role statistical group differences play for the moral acceptability of discrimination. This answer informs the central argument of the dissertation.

1.2 The Central Argument

I argue that when discrimination is rationalized by an appeal to accurate statistical group differences, it frames discrimination as a matter of efficiency and prompts the value of in-group loyalty by making statistical groups socially salient. I furthermore provide an argument for why we should be wary of the implications of such findings, namely because accurate statistics do not alleviate the moral wrongdoing involved in many cases of statistics-based discrimination. Accurate statistics might make us forget that in instances of statistical discrimination, people are (very often) treated as objects or mere containers of statistical probabilities. This is the case when a discriminatory practice reduces people to one or very few of their social or bodily traits and instrumentalizes that reduction to further a goal extraneous to the relational interests of the targeted person(s). In other words, people are reduced and treated as instruments against their immediate interests. To refocus people's

attention away from efficiency considerations and in-group loyalty, we can re-frame statistical discrimination as a practice that wrongs people who are as morally deserving as oneself (making them a perceived in-group) and weaken the in-and-out-group demarcation prompted by the existence of statistical group differences. One way of doing so is to appeal to the moral values that acceptors of discrimination find weighty and thus frame statistical discrimination as a practice that violates the rights of a group of people who are more similar than different to oneself, and who deserve respect for their human dignity.

1.3 Recruitment as a Case

In my dissertation, I focus on recruitment discrimination as a critical test case. I do so for multiple reasons. First, my focus enables theoretical and conceptual precision (see, e.g., Sartori, 1970). The theoretical delineation between taste-based and statistical discrimination stems from economic theories of labor market discrimination. Since I build on and extend these theories, it is natural to focus on recruitment. When the theory of statistical discrimination, for example, stipulates that employers discriminate because they seek to make an economically efficient hiring decision, I critically examine whether there could be other underlying motivations for employers to rely on statistical group differences. My focus on recruitment discrimination enables me to identify gaps and refine the theory with precision without other potentially distracting factors, such as in which sphere of life the discriminatory act occurs. As a result, it is evident how and where I contribute to the existing theories of statistical and taste-based discrimination.

Second, it enables empirical (especially experimental) precision and consistency (see, e.g., Dafoe et al., 2018). I have chosen to investigate the acceptance of recruitment discrimination across all my empirical studies. I have done so to explore whether findings are consistent across studies and contexts but also because the results add to the same story. In other words, by keeping the discrimination sphere constant, I can better speak to how the variables of interest, such as statistical information and moral values, influence the acceptance of discrimination. If I had introduced sphere as another variable of interest, it would be more difficult to ensure information equivalence as sphere could trigger a new layer of imbalanced background assumptions among respondents.

Third, it is what Levy (2008) calls a crucial case. If I do not find that statistical group differences matter in this case, it is less likely that they would play a role in other cases. Recruitment has a clear aim, and group differences may visibly play a role in lieu of adequate information. It is less evident how

credibly statistical information plays a role in other spheres, such as the private (for instance in dating) or political (for instance when voting for specific candidates) spheres. Additionally, the respondents can easily understand the case. Most people have heard about, and some even witnessed, recruitment discrimination. In many ways, recruitment discrimination is one of the most paradigmatic discrimination cases. I am not investigating a complex or convoluted case of discrimination, but a straightforward one. This makes the results more interesting because most respondents are aware that they are assessing discrimination. If people still accept such an evident case of discrimination, they do so with their eyes open.

1.4 Roadmap

In Chapter 2, I present the methodology of the dissertation, and in Chapter 3, I introduce the core concepts and theoretical backdrop. I focus on the concept of statistical discrimination and its classical delineation from taste-based discrimination, both originally formulated as explanatory models of labor market discrimination. Building on this distinction, I present a typology of discrimination that nuances the sharp distinction between statistical and taste-based discrimination and enables me to identify three features of discrimination that could explain the potential moral difference between statistical and taste-based discrimination. The remainder of the dissertation is structured by the three guiding questions. In Chapter 4, I present the empirical methods used to investigate Q1 and the appertaining results. In Chapter 5, I discuss how these empirical insights motivate a moral framework from which we can assess cases of statistical discrimination. I furthermore present the core of the moral framework and reflect on its theoretical and empirical implications, thereby answering Q2. In Chapter 6, I present moral reframing as an experimental method for testing how to influence people's moral assessments, and present and discuss how the experimental results answer Q3. I then turn to Chapter 7 to discuss and reflect on the implications of my dissertation and how the different studies underpin the central argument of the dissertation.

Chapter 2: Methodologies

As evident in the introduction, the research underpinning the dissertation has an empirical and a normative theoretical leg. First, the dissertation investigates whether people assess discrimination as more acceptable when it is rationalized by or reflects statistical differences between groups. The empirical findings prompt a normative question, asking whether it is morally problematic that people accept statistical discrimination. The normative framework and the argument it provides for why statistical discrimination is often morally wrong serve as a justification and motivation for the return to the empirical realm. If we have normative reasons to mitigate people's acceptance of statistical discrimination, it is worthwhile to explore which moral values and appeals are effective in doing so. Thus, the dissertation brings together the methodology of two fields: 1) an empirical political science approach, seeking to identify what affects people's judgments of discrimination and the moral values present in their justifications of their judgments, and 2) an analytical philosophical approach, seeking to assess and analyze the normative principles at stake on the issue of statistical discrimination.

2.1 Experimental (and Empirical) Political Science

I use empirical methods from political science to investigate when and why people accept recruitment discrimination. Whereas I combine different quantitative methods (and a bit of qualitative interpretation), I mainly work experimentally. A crucial part of my dissertation investigates how statistical information and moral arguments influence people's acceptance of discrimination. This entails causal identification. I seek to isolate the effect of narrowly defined and operationalized variables to nuance and contribute to existing theories of recruitment discrimination. The best way to test this kind of causal relationship is via a randomized controlled trial (RCT) (see, e.g., Druckman et al., 2011). Randomization is a way to balance out the different treatment groups so that imbalances in all variables average out to become equivalent across conditions. In turn, we can identify outcome differences between treatment groups and the control group, also known as average treatment effects (ibid.), knowing that the effects are not driven by systematic individual-level differences between the treatment groups. RCTs address selection problems (Angrist & Pischke, 2009; Arceneaux, 2010), such as when people systematically select certain activities, life trajectories, and so forth such that we cannot fully identify whether *what* they selected into or *the reason* they selected into

it in the first place best accounts for the observed effects. RCTs have a strong experimental control because they enable us to design treatments that vary only on the variables of interest and thereby rule out third variable confounders that would likely interfere in an observational design (*ibid.*). In my dissertation, I implement the RCT logic in online vignette-based survey experiments by designing vignettes that strictly vary the variables potentially affecting discrimination acceptance while holding all other information constant (under the assumption of information equivalence. See, e.g., Dafoe et al. [2018]).

Experiments have recently faced much scrutiny and criticism following the identification of systematic malpractice or biases. These include, for example, p-hacking practices, where researchers select specific statistical models (or even subsamples) that produce statistically significant effects (Brodeur et al., 2020; Gerber & Malhotra, 2008), and publication biases (Franco et al., 2014), where journals favor statistically significant results above null findings. These practices have resulted in a replication crisis in the quantitative social sciences (Dreber & Johannesson, 2019). I have therefore implemented open and reproducible science practices (see, e.g., Christensen et al., 2019) such as online preregistration of hypotheses, research designs, and statistical methods before fielding my experiments, and I have conducted sensitivity analyses (Christensen et al., 2019) and conceptual replication when possible (Crandall & Sherman, 2016; Hudson, 2023). In other words, I have approached experimental work and statistical analysis with transparency and reproducibility as core evaluative principles.

Nevertheless, experiments cannot answer all empirical questions—especially not those that try to nuance, explore, and systematize themes in respondents’ opinions and assessments. I therefore also use non-experimental methods such as quantitative text analysis to explore opinion content rather than questions about causality. This reflects the idea of methodological triangulation, where a phenomenon is investigated from different methodological approaches (Larsen, 2021).

2.2 Analytical Philosophy

In the dissertation, analytical political philosophy plays the part neatly described by McDermott (2008): “Whereas social scientists aim to determine the empirical facts about human behavior and institutions, political philosophers aim to determine what ought to be done in light of that information” (p. 11). As such, the insights from the empirical studies of people’s acceptance of discrimination tell us little to nothing about whether we ought to morally accept statistical discrimination to a greater extent than other forms of discrimination. To make such an evaluation, we need to turn to analytical political

philosophy. I further agree with McDermott (2008) that political philosophy and political science have a lot in common. They both seek to uncover truths about our political and social worlds and the implications of such truths for, say, democratic values, norms, and behaviors. The combination of empirical political science and analytical philosophy enables me to uncover how people rationalize behavior that is presumably normatively unwanted and critically examine the implications of such findings; namely, by asking whether it is a moral problem if people accept statistical discrimination.

In the normative part of the dissertation, I use the analytical philosophical method called the reflective equilibrium, which seeks to establish coherence between our considered judgments about morally relevant cases and our general moral principles (Knight, 2017; Rawls, 1999). To pursue the reflective equilibrium (the state of coherence), we must formulate and defend a principle and identify relevant (and analogous) cases on which to critically test the principle. Cath (2016) sketches out the general three-step flow of working with the reflective equilibrium method: First, one holds specific considered judgments about a relevant domain. It could be the judgment that a case of recruitment discrimination is morally wrong. Second, one seeks to come up with a principle that can systemize and account for our initial belief(s). This could be a principle stating that it is morally wrong to discriminate in recruitment practices if it is based on traits irrelevant to the job description. Third, one engages in a reflective process revising the principle or initial judgments by testing whether the principle tracks the judgments about analogous and relevant cases. For example, one could counter an inconsistency between the principle of irrelevance and a judgment about a case where gender is a statistically relevant trait for a job (for instance when gender correlates with cost efficiency). To solve the incoherence, one must revise either the principle or the judgment. Furthermore, the reflective process might get even more complex when we include considerations of relevant background theories (Daniels, 1979). Those could be theories about wrongful discrimination, egalitarian theories, or even (evaluative) theories about what constitutes proper moral judgments.

As evident, it is not an easy task to formulate a plausible normative principle. Importantly, the principle must be explicit about which conditions have to be met to categorize a practice as, say, morally acceptable or unacceptable. Furthermore, the principle should reflect whether its conditions are necessary and/or sufficient wrong-making criteria (Cath, 2016; Sommer Hansen & Midgaard, 2016). In other words, it must be explicit whether, say, irrelevance reflects one of several wrong-making features of recruitment discrimination or whether irrelevance is a necessary requirement for categorizing any case of recruitment discrimination as morally wrong.

In Chapter 5, I follow the three-step flow presented by Cath to formulate a principle of wrongful statistical discrimination.

2.3 The Fruitfulness Approach

A recent discussion on the reliability of laypeople's judgments asks whether these judgments should play any role for the testing and formulation of normative political theories and principles (Andow, 2016; Bengson, 2013; Kumar & May, 2018; Nagel, 2013; Paulo, 2020; Stich & Tobia, 2016). Some scholars argue that lay intuitions should serve as a test of our political theories because we should seek to formulate theories that resonate with laypeople and political agents, and that work properly under the constraints of real-world politics. Other scholars argue that we cannot trust laypeople's intuitions (or sometimes, for that matter, philosopher's considered judgments), and that adapting political theories to empirical intuitions on cases would lead us astray (Copp, 2012; Schwitzgebel & Cushman, 2014; Stich & Tobia, 2016). In this dissertation, my research falls in between these two views. I take an approach termed *the fruitfulness view* (Lindauer, 2020), where political philosophical analysis can provide motivation, guidance, or an evaluation standard for empirical studies and findings. While I believe it is important to consider laypeople's (citizens') moral judgments of discrimination, analytical political theory serves to scrutinize such judgments to test whether we have strong moral arguments or principles against or in favor of the views of laypeople. In the case of statistical discrimination, I argue that we have little reason to believe that the existence of statistical group differences as such makes discriminatory practices categorically more acceptable than they are in their absence, despite laypeople's tendency to assess such cases as more acceptable. In Chapter 5, I engage with the empirical intuitions identified in Chapter 4. Chapter 6 strongly reflects the fruitfulness view, as I test the constraints and methods that would make moral arguments fruitful or effective for decreasing people's acceptance of statistical discrimination. Thus, whereas normative discrimination scholars have certain theoretical ideas and principles guiding their moral assessments of statistical discrimination, Chapter 6 demonstrates that some of the most common principles (namely equality, fairness, and individuality) might produce unwanted effects such as increased political polarization and even increased acceptance of statistical discrimination among more conservative-leaning individuals.

Whereas the methodology of each article falls into either analytical political philosophy (ideas as objects) or empirical political science (observations as objects), the overall research project, and thus the dissertation, resembles the methodology used in the field of experimental philosophy, which broadly

defined is an iterative process wherein philosophical theories, ideas, and principles are tested or compared to the principles, intuitions, and arguments empirically identifiable amongst laypeople (Stich & Tobia, 2016).

Chapter 3: Theories of Discrimination

Discrimination is not *one* thing. Discriminatory practices vary in who constitutes the target of discrimination, in what spheres of life they occur, and whether discrimination is direct or the result of disparate outcomes across groups (Bertrand & Duflo, 2017; Lippert-Rasmussen, 2018). They also vary on what underpins the process of drawing distinctions between groups in the first place—do people draw distinctions between groups on a whim or because of animus, inaccurate beliefs, or statistically identifiable differences? In this dissertation, I examine the latter motivation as I investigate the role of statistical group differences in assessing the moral acceptability of recruitment discrimination.

The point of departure is the classical distinction from the economics of labor market discrimination. This is the distinction between taste-based discrimination and statistical discrimination, where the former describes labor market discrimination that occurs (in the short term) when employers are willing to sacrifice profits to avoid association and transactions with a specific population group (Becker, 1957). Such discriminatory decisions are rooted in distaste or even animus against the discriminated group. Statistical discrimination, by contrast, seeks to explain long-term labor market discrimination as it (in many instances) is an economically efficient way to decide between applicants. The theory of statistical discrimination was formulated by Arrow (1973) and Phelps (1972) and stipulates that employers discriminate by drawing on statistical group differences when assessing the fitness of candidates. This happens when employers lack adequate information about the applicants, and a group's statistics become proxies for the individual's fitness, efficiency, or performance. In short, statistical discrimination occurs when an agent seeks to obtain a (presumably legitimate) goal and does so by discriminating against a social group because the agent is of the belief or assumption that doing so will instrumentally further the goal (Schauer, 2018).

This distinction between taste-based and statistical discrimination illustrates that discrimination can stem from an illegitimate goal (e.g., animus) and be cost-inefficient (as is often the case with taste-based discrimination), which might fit the general assumption or perception of discrimination. However, it also illustrates that discrimination can stem from the pursuit of a legitimate goal and be cost-efficient. Indeed, studies have shown that people tend to assess statistical discrimination as more rational (Khan & Lambert, 2001) and morally acceptable (Tilcsik, 2021) than taste-based discrimination.

3.1 A Typology of Discrimination

I have identified three parameters on which taste-based and statistical discrimination vary. First, they vary on what motivates the employer to discriminate in the first place. Employers discriminate either because of distaste or because of a cost-efficiency goal. Second, they vary on whether they rely on statistical group differences. Third, in cases where employers rely on statistical group differences, we can distinguish between whether these are statistically accurate or inaccurate (Bohren et al., 2023). As a result, I have developed a typology of discrimination that nuances the differences between taste-based and statistical discrimination. The typology is illustrated in Table 1.

Table 1. Typology of Discrimination

	Intention	Statistical indicator	Accuracy
Taste-based	Discriminatory	No	-
Biased statistical	Discriminatory	Yes	Accurate
Accurate statistical	Non-discriminatory	Yes	Accurate
Inaccurate statistical	Non-discriminatory	Yes	Inaccurate

The typology illustrates that we can conceive of cases of discrimination where employers rely on statistical group differences in their decision-making *because* they have distaste for specific groups. Independent of the employer’s motivation, the statistical group differences drawn upon can be either accurate statistical proxies (and thus efficient predictors) or inaccurate statistical proxies (inefficient and spurious). To find out what about statistical discrimination might influence its moral acceptability, it is necessary to conceptually and experimentally pry apart the three morally relevant features identified above. Henceforth, I propose to think of biased, accurate, and inaccurate statistical discrimination as *statistics-based* discrimination, as only accurate statistical discrimination matches the original theoretical conceptualization of *statistical* discrimination.

3.2 The Moral Relevance of Intention, Statistics, and Accuracy

The three identified features in Table 1 further reflect factors that could affect people’s moral assessments of discrimination. This is because the features tap into some of the most prominent moral objections against discrimination.

With regard to intentions, various studies of moral judgments find that intentions influence people's moral assessments, and intentions also play an integral role in law, for instance delineating manslaughter from murder (Cushman, 2015; Greene et al., 2009; Young & Saxe, 2011). In the normative literature on discrimination, a prominent set of accounts (called the mental-state accounts) stipulate that discrimination is wrongful when it stems from a specific set of intentions, dispositions, or motivations (Lippert-Rasmussen, 2013). An example of a prominent account is Eidelson's (2015) disrespect account, stipulating that discrimination is morally wrong when it stems from attitudes of disrespect or contempt (or otherwise manifests disrespect). This suggests that the employer's intentions (whether based on distaste or cost efficiency) could determine or at least influence people's acceptance of statistical discrimination.

There are likewise reasons to expect that the reliance on a statistical proxy plays a role in people's acceptance of statistical discrimination. This might be because reliance on statistical group differences appears to be a more objective way to assess applicants than when assessment is based on personal beliefs or idiosyncratic preferences. Indeed, some persuasion studies find that appealing to facts or scientific evidence influences people's normative and political opinions (Altonji & Pierret, 2001; Bullock, 2011; Gilens, 2001; Sides, 2016). Furthermore, statistical group differences on, say, job performance, cost efficiency, or longevity of employment make social group membership a (at least) statistically relevant trait. Relevancy of traits is another prominent discussion in the normative literature on discrimination as some scholars argue that discrimination is morally wrong if it is based on irrelevant traits (Halldenus 2018). Such considerations are common in anti-discrimination laws such as the American federal law prohibiting gender-based hiring discrimination unless gender qualifies as a bona fide occupational qualification (Title VII of the Civil Rights Act of 1964). It might be that people perceive statistical group differences to be a feature that makes social groups a relevant consideration and, therefore, more morally acceptable than when discrimination does not rely on a statistical proxy.

However, the moral relevance of statistical proxies might be tied closely to whether they are accurate or inaccurate. This is because accurate proxies are more likely to effectively obtain the legitimate goal (for instance hiring the best applicant) and thus produce some benefits, whereas inaccurate proxies are less likely to produce the intended goal and, thus, more likely to create harm. The accuracy of the statistical proxy might therefore reflect the amount of harm caused by the discriminatory practice. The moral relevance of harm is prominent in the normative literature (Arneson, 2017; Berndt Rasmussen, 2019; Lippert-Rasmussen, 2013) and is a feature that influences people's

assessments of moral cases and dilemmas (Barak-Corren et al., 2018; Cushman et al., 2006; Gray et al., 2022) such as cases of discrimination (Albertsen et al., 2023).

As discussed above, we have empirical evidence suggesting that intentions, statistical group differences, and the accuracy of the statistical indicator influence people's moral assessments, and we have normative, theoretical arguments for why this might be—namely, because the features reflect normatively significant differences in the motivation, relevance, and consequences of discriminatory actions. In sum, the dissertation builds on a typology of discrimination (Table 1) to explore what features of statistical discrimination might drive people to accept it, and whether such features are morally significant in that they *should* influence our theories of wrongful discrimination.

Chapter 4: When and Why do People Accept Recruitment Discrimination?

As mentioned, some studies find that people tend to accept statistical discrimination more than non-statistical discrimination (such as taste-based discrimination). However, as pointed out in the previous chapter, we do not know exactly what it is about statistical discrimination that would drive this increased acceptance, as the existing studies of people's acceptance of discrimination (Bunel & Tovar, 2021; Khan & Lambert, 2001; Tilcsik, 2021) do not isolate the individual features of discrimination. Maybe people would not accept statistical discrimination if it stemmed from (perceived) illegitimate intentions or inaccurate statistical group differences. This chapter explores when and why people accept recruitment discrimination (Q1).

The chapter is structured as follows. First, I present the experimental and statistical methods used to (1) identify the causal relationship between the three identified features of discrimination (intention, statistical indicator, and accuracy) and people's acceptance and (2) establish the correlational evidence of the relationship between acceptance status and moral themes in people's justifications. Second, I present and discuss the results, and third, I reflect on the implications of the results for the dissertation's research question.

4.1 Methods and Research Design

I used an experimental approach to investigate whether people accept statistical discrimination to a greater extent than taste-based discrimination and which feature drives people's acceptance. As mentioned in Chapter 2, experiments enable me to leverage randomization to test whether people's average acceptance of discrimination differs across vignette treatments, varying on the identified features of interest. The dependent variable is therefore discrimination type. All the vignettes contained information about a public leader who had to hire a new employee and decided to reject all applicants with immigrant/Hispanic backgrounds. The vignettes differed in whether the public leader was motivated by efficiency considerations or animus towards the rejected group and whether the latter group statistically differed from the general population in their likelihood of accepting a bribe.

I conducted two studies with a between-subjects design, one with Danish respondents (Study 1) and one with American respondents (Study 2). The two studies' treatments differed slightly. In Study 1, the emphasis was on the

employer rationales and the vignettes were based on realistic narratives, describing the employer's intentions and deliberation process. Study 2 utilized a stricter experimental design, cutting back on the thick descriptions, to clearly isolate the effect of intention and appeals to statistical group differences on respondents' moral acceptance. Both studies were preregistered before being fielded. The differences between the studies enabled me to investigate whether people's acceptance is dependent on the employer's individual rationalization/deliberation process or whether the effects hold up when the accuracy of group differences is presented as a fact external to the employer's deliberation process. Both studies aimed at providing equivalent background information across the experimental conditions (Dafoe et al., 2018). The difference in national context between the two studies further enabled me to reflect on whether the results could be generalizable across national contexts.

I included three dependent variables in both studies. To answer what feature of discrimination drives people's acceptance, I included moral acceptance as the main dependent variable. After reading the vignette, respondents were asked to indicate their degree of moral acceptance on a seven-point Likert scale, running from completely unacceptable (1) to completely acceptable (7). To examine the role of moral values for people's acceptance, Study 1 included an open-ended question, asking the respondents what their reasons were for their assessment, and both studies included a question asking respondents to what extent they agree or disagree that six different moral principles were essential to their assessments (on a seven-point Likert scale). The moral principles concerned employer intentions, disrespect towards the discriminated applicants, the harm resulting from the decision, the benefit resulting from the decision, the relevance of the group trait (ethnicity), and whether people were treated as individuals. I used OLS regression to analyze the results, comparing the average acceptance or agreement means across treatment groups.

To test whether the moral themes in people's open-ended justifications differed across those accepting and opposing discrimination, I used structural topic modelling (Roberts et al., 2014) to identify which themes emerged in the justifications and whether they correlate with acceptance status. I furthermore used a keyness analysis (Benoit et al., 2018) to investigate whether acceptors of discrimination appeal to distinct moral values compared to opposers of discrimination. In total, this chapter examines and summarizes the following three relationships:

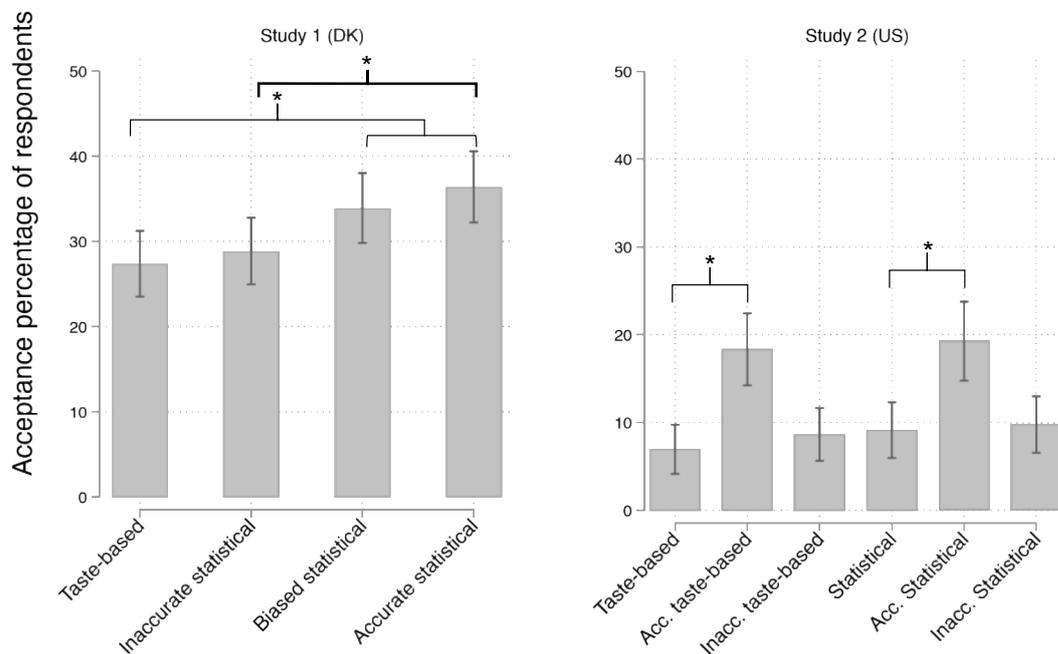
1. The causal influence of discrimination type on people's moral acceptance.
2. The causal influence of discrimination type on people's agreement with moral principles.

3. The correlational relationship between acceptance status and moral values in justifications.

4.2 Results

The experimental results provide evidence for whether all types of statistics-based discrimination are assessed as more acceptable than taste-based discrimination and, relatedly, what feature of discrimination drives people’s acceptance. The results are depicted in Figure 1, showing the percentage of respondents who accept (or tolerate) discrimination.

Figure 1. Acceptance across Discrimination Types



Note. The bars indicate the percentage of respondents who accept or tolerate (respondents indicating a neutral stance) discrimination across conditions. The error bars indicate the 95% confidence intervals estimated with robust standard errors. * indicates that the estimated difference between the compared groups is statistically significant at $p < 0.05$. Study 1 $n = 2055$. Study 2 $n = 1939$.

The results depicted in Figure 1 show that people do not categorically accept all types of statistics-based discrimination more than taste-based discrimination; however, they do accept accurate statistical discrimination more than taste-based discrimination. Meanwhile, when discrimination is based on legitimate intentions *but also* on statistically inaccurate group differences, people assess it as equally (un)acceptable as taste-based discrimination. The results are robust when regressing discrimination type on the original Likert-scaled moral acceptance variable. It is worth mentioning that in both studies, the majority of respondents oppose the employer’s discriminatory decision.

Next, we turn to the feature of discrimination that drives people's acceptance. The results strongly suggest that the accuracy of the statistical indicator influences people's acceptance most starkly. In other words, people care about whether there in fact exist statistical differences between the discriminated group and a relevant comparator (such as the general population) more than they care about the intention of the employer or even the reliance on a statistical indicator in itself. In Study 1, nearly 24% more respondents accept biased statistical discrimination compared to taste-based discrimination ($p = 0.023$), and nearly 33% more respondents accept accurate statistical discrimination compared to taste-based discrimination ($p = 0.002$). Approximately 26% more respondents accept accurate statistical discrimination compared to inaccurate statistical discrimination ($p = 0.01$).

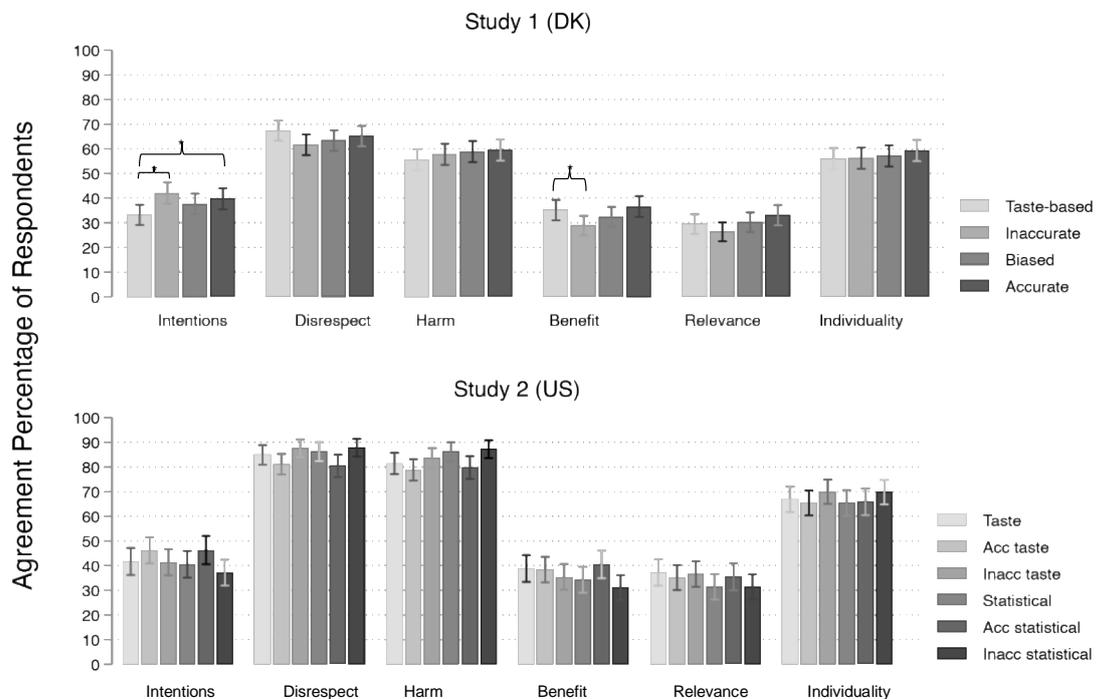
The same pattern unfolds in Study 2. Compared to the taste-based control condition, 165% more respondents ($p = 0.00$) accept the accurate taste-based condition (where respondents are informed about statistical group differences between the discriminated group and the general population). For the statistical conditions, 111% more respondents accept accurate statistical discrimination compared to the statistical discrimination control condition ($p = 0.00$). Like in Study 1, the accuracy of the statistical group differences prompts people to accept discrimination, rather than the employer's intention or the employer's belief in statistical group differences.

These findings provide a partial answer to Q1. Namely, discrimination is perceived as more morally acceptable *only when* it is based on perceived statistically accurate group differences. This means that even when employers are motivated by distaste or animus, people are more willing to accept the discriminatory action if the rejected group statistically differs from the general population. Whereas the effects identified across the two studies are similar in direction, it is worth noting that (1) many more Danish respondents accept all types of discrimination compared to the American respondents. This suggests that the effect might cut across national contexts, irrespective of the population's initial willingness to accept discriminatory practices. It is also worth noting that (2) among Danes, more respondents are swayed from a neutral stance (the middle category, "neither acceptable nor unacceptable") towards active acceptance, whereas for Americans more respondents are swayed from active opposition towards the neutral stance. This connects well with the descriptive finding that opposition is generally more widespread among the American respondents compared to the Danish respondents.

Next, I turn to the moral values people find essential for their assessments. First, I examined the moral principles with which respondents were asked to indicate agreement. If they differ across treatment conditions, it could indicate that different moral values or principles play a role for people's

acceptance of discrimination based on statistically sound group differences. The results of these analyses are not reported in Articles A and C but are based on the data from the two experiments. The results are shown in Figure 2.

Figure 2. Agreement with the Importance of Various Moral Principles for Assessment



Note. The bars indicate the percentage of respondents who agreed that the principle was essential in their moral assessment. The error bars indicate the 95% confidence intervals estimated with robust standard errors. * indicates that the estimated difference between the compared groups is statistically significant at $p < 0.05$. Study 1 $n = 2055$. Study 2 $n = 1939$.

As evident in Figure 2, there are no stark differences in which principles people find important depending on discrimination type. Rather, the results show that people generally indicate that they assessed the employer's decision based on whether the action signaled disrespect toward the discriminated group, whether it harmed the discriminated group, and whether applicants were treated as individuals. As Figure 2 also shows, there are a few statistically significant differences across discrimination type. In Study 1, the percentage of Danish respondents agreeing that intentions were important is 6.5 percentage points greater in the accurate statistical ($p = 0.03$) condition and 9 percentage points greater in the inaccurate statistical ($p = 0.003$) condition, compared to taste-based discrimination. This could simply indicate that they recognize the lack of malice in the employer's intention, regardless of the accuracy of the statistical indicator. There were also 6.3 percentage points fewer respondents who agreed that benefit was important in the inaccurate statistical condition compared to the taste-based condition ($p = 0.031$). This may reflect that

respondents recognize that inaccurate proxies will likely not prompt recruitment decisions that benefit the workplace. In Study 2 there were no statistically significant differences, but 6 percentage points fewer American respondents appeared to find considerations of harm important in the accurate statistical condition compared to the statistical control condition ($p = 0.053$). However, these differences offer no clear evidence of whether certain moral principles could play a role in why accurate statistical group differences increase people's acceptance. Moreover, it forces respondents to think of their assessments in relation to these very specific moral values instead of enabling them to freely formulate what mattered to them. I therefore turn to analyzing the relationship between acceptance status and moral values that emerged in the respondents' open-ended justifications.

Figure 3 shows the results of a structural topic model (for the top six topics among opposers and acceptors, respectively) for the Danish sample, testing whether acceptance status (accepting or opposing discrimination) predicts which themes are prevalent in the respondents' justifications.

The results indicate that acceptors and opposers largely rely on different topics in their justifications, and a deep dive into the top words and responses most highly associated with each topic shows that opposers justify their stance by appealing to the badness of racism, discrimination, breach of meritocratic principles, generalizations, and the importance of equality.

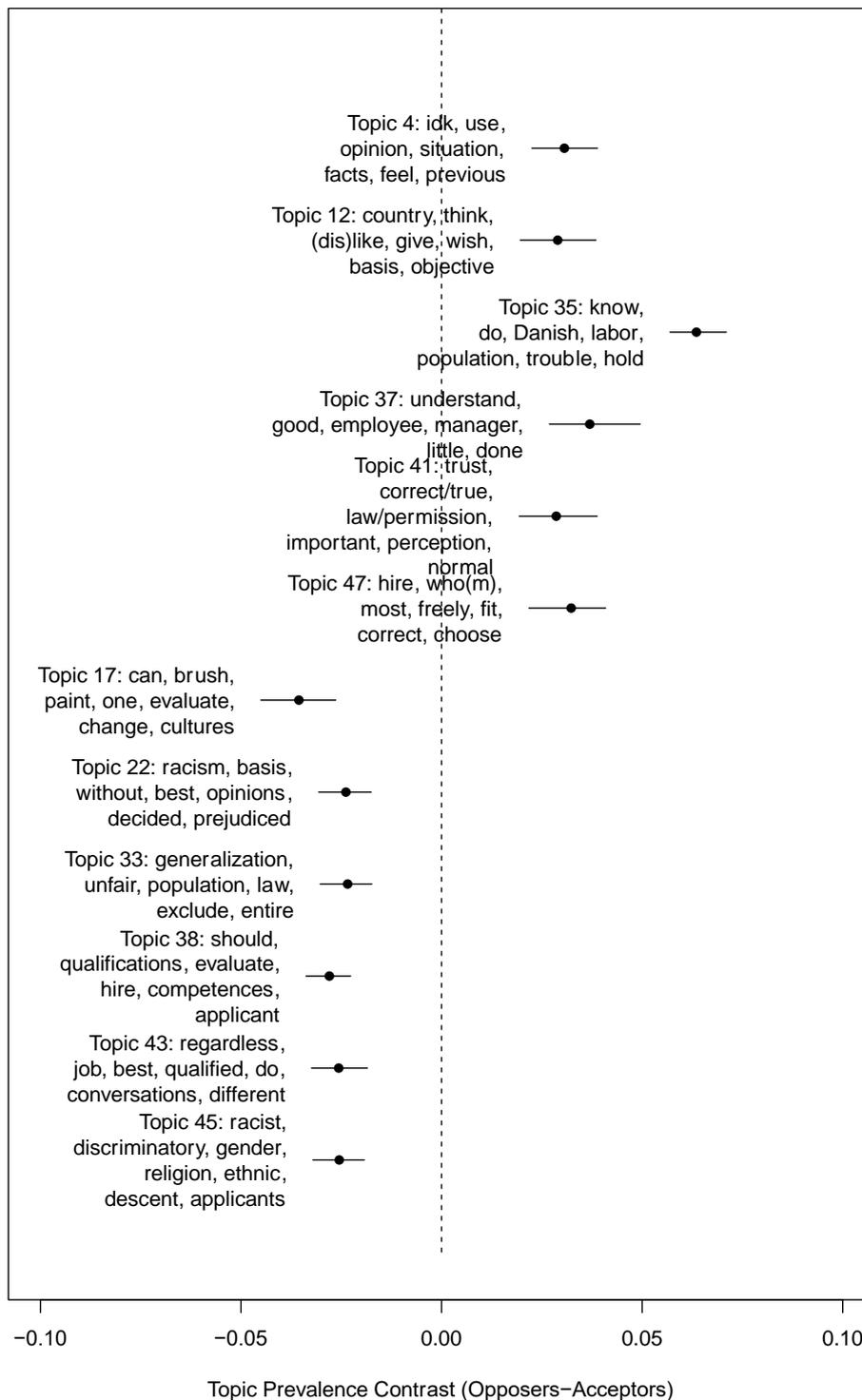
Anonymous Opposing Respondent [translated]: “[The employer’s] decision is somewhat culturally racist, because when it comes to hiring employees, there should be an individual assessment of the employee rather than evaluating them based on their culture, gender, ethnicity, or religion”. (Response highly associated with Topic 45).

In contrast, acceptors of discrimination appeal to themes such as the complexity of recruitment decisions, the importance of in-group loyalty, employer's rights and status, and using statistical group differences as objective evidence.

Anonymous Accepting Respondent 1 [translated]: “[The employer] bases his approach on facts. He doesn't want to be in the same situation as before. Cultural differences often become a big problem.” (Response highly associated with Topic 4)

Anonymous Accepting Respondent 2 [translated]: “If, as a manager, he believes that the best work environment he can and wants to lead is without foreigners, it is his right and duty as a manager.” (Response highly associated with Topic 12)

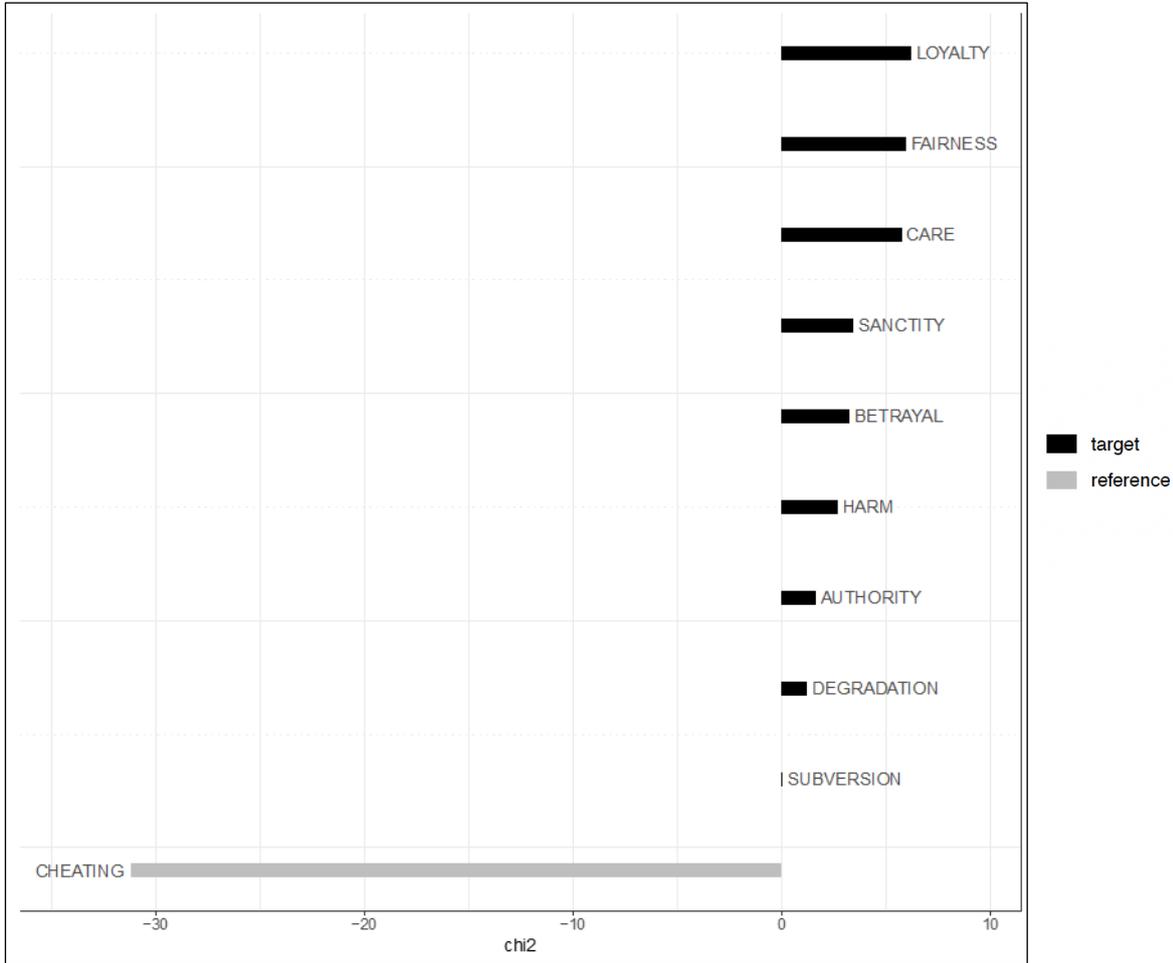
Figure 3. Difference in Topic Prevalence across Acceptance Status



Note. The figure shows the six topics most strongly correlated with acceptors and opposers, respectively, and the FREX words for each topic. The estimates are the mean difference in topic proportions between acceptors and opposers. Positive values indicate a higher probability of prevalence among acceptors, and negative values indicate a higher probability of prevalence among opposers. The dotted reference line indicates the value where a topic is equally prevalent among acceptors and opposers.

When the responses are filtered by the Moral Political Dictionary (Jung, 2020; and translated into Danish by Simonsen & Widmann, 2023) so that only moralized words are included (1,827 responses and 25,578 words), the results echo the themes identified with the structural topic model. Opposers use more moral words related to the vice of cheating than acceptors ($p = 0.000$), and their most frequent vice words are *discrimination, prejudice, differential treatment, discriminatory, and discriminate*. Acceptors instead use moral words related to the virtues of loyalty ($p = 0.013$), fairness ($p = 0.015$), and care ($p = 0.016$). Among acceptors, the most frequent loyalty words are *belief, reliable, trustworthy, and loyal*. The top fairness words are *fair, honest, reasonable, proper, and honesty*. The top care words are *protect, worry, safe, and guard*. The results are shown in Figure 4.

Figure 4. Keyness Statistics for Moral Foundations



Note. The relative frequency of moral foundations reflected in acceptors’ justifications of discrimination relative to opposers. Target = Acceptors and reference = Opposers.

The results from the structural topic model and keyness analysis suggest that acceptors and opposers rely on different moral values and principles when assessing recruitment discrimination and, thus, that moral values play a role in people's acceptance of (statistical) discrimination. Those willing to accept discrimination care a lot about workplace loyalty, and statistical differences likely accentuate differences that, in turn, might prompt cultural and ethnic in-group loyalty and favoritism. In other words, statistics-based discrimination might prompt ethnocentric attitudes (Borinca et al., 2023; Evans & Krueger, 2009; Kinder & Kam, 2010) *because* it makes group differences statistically relevant and thus salient.

4.3 Reflections on the Research Question

The results of Studies 1 and 2 provide an answer to when and why people accept recruitment discrimination (Q1). The answer is twofold. First, both Danish and American respondents are more likely to accept recruitment discrimination when it reflects or can be rationalized by statistical group differences. However, that is irrespective of whether the discriminatory action is motivated by animus or efficiency considerations. Second, the analysis of the open-ended responses shows that those who accept discrimination focus distinctly on perceived in-group and out-group differences, such as cultural and moral differences, and that they perceive the discriminated group as an out-group that is statistically and, therefore, morally different from the perceived in-group. Thus, information about statistical group differences makes social groups more salient, prompting people to perceive the discriminated group as a threat to the workplace, current employees, and the perceived in-group.

As such, these findings give a partial answer to the core research question of the dissertation, asking *what role statistical group differences play in the moral acceptability of discrimination*. They play the role of emphasizing group differences, making such differences salient, and, thus, prompting people to evaluate discrimination based on the values of efficiency, objectivity, and (in-group) loyalty.

Chapter 5: When and Why is Statistical Discrimination Morally Wrong?

The findings in Chapter 4 illustrate that approximately one-third of Danes and one-fifth of Americans are willing to accept discrimination when it reflects or is rationalized by accurate statistical group differences. This motivates a normative question. Could the 20-35% of people who accept such discrimination be morally correct in their assessments? Do accurate statistical group differences make discrimination morally acceptable? In this chapter, I dive into a normative analysis of the morality of accurate statistical discrimination by answering the question of *what makes accurate statistical discrimination morally wrong when it is* (Q2). This part of the dissertation uses an analytical philosophical approach (specifically the reflective equilibrium method presented in Chapter 2). I thus move reflectively between considered case judgments and principle formulation to finally propose a moral framework for assessing the acceptability of statistical discrimination.

Chapter 5 has the following structure: First, I give a brief overview of some prominent moral accounts of what makes statistical discrimination wrong, when it is. Second, I present the central test case that accounts of wrongful statistical discrimination should be able to explain. Third, I present my moral framework, developed to assess cases of accurate statistical discrimination. This framework identifies statistical discrimination as a pro tanto wrongdoing when and if it wrongfully objectifies those discriminated against. Fourth, I discuss how the framework relates to the findings presented in Chapter 4, and finally, I reflect on how my moral framework helps us answer the dissertation's core research question.

5.1. Prominent Objections to Statistical Discrimination

Most normative accounts of discrimination pertain to discriminatory actions broadly speaking, implying that most, if not all, discriminatory actions are wrong (when they are) because they entail the same wrong-making feature. Some of the most influential overarching accounts are the mental-state accounts (Alexander, 1992, 2016; Arneson, 2006), the disrespect accounts (Eidelson, 2013; Hellman, 2008), the harm-based accounts (Arneson, 2017; Lippert-Rasmussen, 2013), and the irrelevance accounts (Halldenius, 2018). However, as also reflected in Chapter 4's findings, people do not morally judge all types of discrimination as equally wrong (Figure 1), and they find several

moral principles essential in their assessments (Figure 2), rather than a single principle.

I argue that it is unlikely that one single account with one central wrong-making feature (such as harm, disrespect, or irrelevance) can explain *every single instance* of wrongful discrimination. This is because practices that amount to accurate statistical discrimination are very different from acts of discrimination that are based on distaste or animus and that have little chance of benefiting anyone apart from those who hold such discriminatory preferences. Accurate statistical discrimination, on the other hand, seeks to accomplish a legitimate goal (Maitzen, 1991; Schauer, 2018), might produce benefits for the workplace or even society as a whole, and is assumed to stem from non-discriminatory mental states or preferences (ibid). Since accurate statistical discrimination has this very distinct structure, I argue that we are better off pursuing a theory of wrongful (accurate) statistical discrimination rather than a theory of wrongful discrimination broadly speaking.

I am not the first, however, to formulate an account that explains the wrongfulness in cases of statistical discrimination. A family of normative accounts already seeks to identify and explain what is morally wrong with statistical discrimination, when it is. These are the individuality accounts, which locate the wrong-making feature in whether statistical discriminatory practices fail to treat the people they deal with as individuals. Each of these accounts propose different criteria for what it requires to treat a person as an individual. Eidelson (2015) stipulates that we could fail to treat someone as an individual either if we do not recognize their autonomous capacities or if we do not recognize their interests as equal to other people's interests. Lippert-Rasmussen (2011) proposes that we might fail to treat people as individuals if we dismiss or fail to include reasonably available information about the person we treat. That includes both statistical and non-statistical information. Lastly, Beeghly (2018) discusses an account that stipulates that we fail to treat people as individuals when we treat them as tokens of a type.¹ This entails seeing people as substitutable with other tokens of the same type.

I argue that these accounts do not provide a moral framework to adequately assess when and why (accurate) statistical discrimination is morally wrong. I demonstrate why by testing the accounts on a case of statistical discrimination that I believe we would want to categorize as a moral wrongdoing. I further argue that the accounts pick up on very relevant features, or symptoms, of what makes statistical discrimination wrong when it is, but that we need to understand how all the features, or symptoms, relate to each other.

¹ Beeghly does not think that this account successfully explains what is wrong with stereotyping (which also faces the objection that it fails to treat people as individuals).

Together, they reflect a wrongdoing deeper than one of being treated as an individual: They reflect that statistical discrimination is morally wrong if it wrongfully objectifies a person.

5.2 Unchangeable by Will

As mentioned, I formulate a case central to testing the adequateness of accounts of wrongful statistical discrimination. I call the case *Unchangeable by Will*, and it entails the following:

Unchangeable by Will: Jasmin, a highly talented and exceptionally skilled applicant, is applying for a PhD fellowship. Academically, she is an obvious match. The graduate school carefully considers all candidates as the fellowship is prestigious and highly sought after. They find out that Jasmin has a degenerative disease that most likely will cause physiological deterioration and estimate that she would be more economically costly than other potential PhD fellows because of the potential need for assistive technologies. They also estimate that she would be more costly because she is less likely to complete her dissertation within the standard time limit. The graduate school decides to reject Jasmin because of her disease (disability) as it is likely to make her more costly than other applicants. In this case, Jasmin is discriminated against (treated disadvantageously) based on her disability, which functions as a statistical indicator that the graduate school relies on to reach the legitimate goal of hiring the most cost-efficient skilled applicant (Article B, p. 4).

By discussing the potential of each of the individuality accounts for explaining the wrongdoing in *Unchangeable by Will*, I demonstrate that if the accounts stand alone, they cannot adequately explain how (or even identify that) Jasmin is morally wronged. I argue that Eidelson's autonomy account fails because Jasmin should not have to prove to the graduate school that her disease would not make her a costly PhD student. This inflicts a burden on Jasmin and signals that what matters is whether she can, through autonomous will, work twice as hard as others to compensate for her disease. Furthermore, I argue that Eidelson's interest account also fails to identify the case as a moral wrongdoing as the graduate school weights every applicant's interests equally against their own interest: to select the best and most cost-efficient PhD candidate for the program. Lippert-Rasmussen's information account is also insufficient on its own because the graduate school considers Jasmin's skills and qualifications, but her disease crowds out all other considerations and determines the graduate school's decision. Last, the token account (Beeghly, 2018) comes up short because Jasmin is not treated as *merely* a token of a type. The graduate school does not treat her as if she were interchangeable with any other person with a disability; rather they estimate *her individual* chances of

being costly due to her disease, and in her case, performance statistics about people with disabilities determine their assessment of *her as an individual*. Again, the graduate school does not ignore all other information about Jasmin, but the estimation that she will probably be costly hinges on the socially salient group trait she holds (her disease).

5.3 Criteria for Wrongful Objectification

Instead of trying to root the wrongfulness of statistical discrimination in individuality concerns, I propose that statistical discrimination constitutes a pro tanto moral wrong-doing if it wrongfully objectifies someone. This is because whereas being treated as a person sometimes entails being an individual and sometimes being part of (and recognized by) a group, it never entails being merely an object. In short, a person is never reducible to mere objecthood. However, when people are crudely reduced and treated as mere probability estimates against their relational interests, I argue that people are wrongfully reduced to the status of an object. Whereas other conceptions of objectification imply that objectification can be both morally acceptable and unacceptable, my conception concerns cases that are always pro tanto wrongful. This is because people are objectified on three central features of personhood—namely, that their human complexity is ignored, they are perceived as an instrument rather than an end, and their (relational) interests are dismissed. The formal definition of wrongful objectification is thus:

X wrongfully objectifies Y if X reduces Y to one or a few of her properties, P, (the reduction criterion) to use P as an instrument that furthers a goal (the instrumentalization criterion) extraneous to Y's interaction-specific interests (the extraneity criterion) (Article B, p. 15).

We can return to *Unchangeable by Will* to better understand how wrongful objectification identifies and explains what is morally wrong in cases of statistical discrimination. First, the reduction criterion entails that a person is being reduced to less than what they are in their entirety. They are split into parts that serve a certain function or goal (Andrighetto et al., 2017; Gervais, 2013). That reflects part of how the graduate school treats Jasmin; Jasmin has many facets, traits, and skills, but to find the most cost-efficient applicant, the graduate school reduces her to her disease. Second, instrumentalization entails the Kantian notion that a person is treated as an instrument to reach someone else's goal (Nussbaum, 1995). Statistical discrimination is essentially a form of instrumental discrimination (Schauer, 2006, 2018), and, in Jasmin's case, the graduate school uses her disease as an instrumental proxy, enabling them to estimate whether rejection or acceptance will promote their recruitment goal. Last, the extraneity criterion must be met before we can categorize

something as wrongful objectification. This criterion entails that a person's immediate and relational interests are dismissed or ignored. Interests are central to the concept of objectification because it is counterintuitive to claim that one is treated as an object while having one's interests acknowledged and promoted. This is especially true in the case of the interests at play in interactions between people as these are personal and relational rather than abstract and general. In *Unchangeable by Will*, Jasmin signals an immediate and interaction-specific interest in becoming a PhD student by applying to the graduate school (an abstract and general interest could be to have a functioning meritocracy, but that is not a relational and interaction-specific interest). However, the graduate school's discriminatory process pursues an end *extraneous* to Jasmin's interaction-specific interests. When all three criteria are met, we are dealing with a case of wrongful objectification, where a person's complexity, status as an end, and relational interests are ignored.

5.4 The Framework's Scope

Whereas I use *Unchangeable by Will* as a crucial test case and an illustrative example of how the normative framework makes sense of cases of statistical discrimination, I also analyze several other (more paradigmatic) cases of statistical discrimination such as racial profiling and insurance pricing. These analyses and their normative conclusions reflect that many cases of statistical discrimination entail wrongful objectification. This may cause concern for some readers. Is the theoretical framework overinclusive? I argue that it might at first glance seem so, but that is not the case.

First, it is important to pay attention to what it means to be severely reduced. I build on a broadly acknowledged understanding of objectification as a reduction that entails that one is severely reduced to a certain function or bodily traits. This includes when people are reduced to their skin color, height, weight (and other bodily traits), or their bodily or social function. Examples of the latter could be the bodily function of walking (or not walking) or the social function of being a mother or a Muslim. The reduction must pertain to one's bodily traits or socially salient function. This means that many forms of reduction are not of the sort that falls under the reduction criterion. These could include instances when one is reduced to owning a specific brand of bike, baking a lot of cakes, only wearing blue hair clips, and so forth. Being reduced to such facts is not a reduction to a bodily trait or function or a social trait.

Second, I believe that a framework should not be biased in the sense that it only identifies cases as wrongful if they pertain to a group we care for. As such, one might think that some social functions come with privileges and, thus, wrongfully objectifying discrimination against such people is morally

permissible. Is it really morally bad if, say, CEOs are statistically discriminated against? In short: Yes, it is bad because they too are people, not objects. However, there is a caveat to this conclusion, leading me to the third reason why the framework is not overinclusive.

Third, the framework stipulates that wrongful objectification is a *pro tanto* wrong, leaving it open that certain circumstances could justify wrongful objectification and, thus, cases of statistical discrimination. In some instances, there could be other considerations of justice that render statistical discrimination morally permissible. This might be in cases where statistical discrimination is necessary to mitigate severe social or historical injustices, or when the consequences of refraining from statistical discrimination would be too grave. This, however, does not mean that the objectified person did not face the moral costs of such treatment, as the subjective experience of being wronged presumably still lingers (see, e.g., Da Silva, 2021). It instead means that the objectifying action is *overall* permissible because weighty justificatory reasons or circumstances apply. Such justificatory reasons or circumstances should be frequently and thoroughly scrutinized to make sure that wrongfully objectifying discriminatory practices do not exist without being inherently necessary to justice.

Taken together, the objectification framework I propose does not uncritically categorize every action of statistical discrimination as morally impermissible. Rather, it forces us to reflect on important questions about the discriminatory action in question: In what way does the action constitute a reduction, and to what? Are our judgments about discrimination cases colored by an unwillingness to grant respect of personhood to certain groups or people? Are we faced with a moral dilemma, and if so, are there weighty justificatory reasons for why we should not refrain from engaging in a statistically discriminatory practice?

5.5 Reflections on the Research Question and Implications

The normative analysis and framework help me answer the second guiding question: *What makes accurate statistical discrimination morally wrong when it is?* (Q2). Statistical discrimination constitutes a *pro tanto* moral wrongdoing when it reduces people to one or a few of their bodily or social traits in order to instrumentally pursue a goal that is extraneous to one's immediate and relational interests. In other words, it is morally wrong when and because it wrongfully objectifies people.

When using the wrongful objectification framework to discern which cases of statistical discrimination are morally wrong and why, it is evident that many

cases of statistical discrimination entail wrongful objectification. As such, it becomes pressing to scrutinize statistically discriminatory practices rather than blankly permitting or condoning them simply because they rely on accurate proxies. Considering the results presented in Chapter 4, I believe it is reasonable to explore methods to mitigate the acceptance of wrongfully objectifying practices of statistical discrimination. In the case of recruitment discrimination, the action, the discriminatory outcome, and the disadvantage imposed on the discriminated group remain the same regardless of whether the employer relies on (accurate) statistical beliefs. Applicants are rejected solely because of their ethnic group membership, and it is far from evident that such treatment is permissible with reference to other weightier requirements of justice. Coupled with Chapter 4's findings regarding people's justifications of discrimination, acceptance seems to reflect a demarcation of in- and out-groups rather than providing arguments for why, all things considered, statistical discrimination is a necessary tool for obtaining justice.

As such, people may be more willing to ignore requirements of treating applicants as persons if they are perceived as belonging to the out-group. After all, objectification and dehumanization often target perceived (ethnic) out-groups (Gervais, 2013; Vaes et al., 2012), and one of the drivers of such dehumanization and objectification is stereotyping (Kronfeldner, 2021)—a process very likely prompted by perceived statistical demarcations between groups. If this holds, mitigating or reducing acceptance of statistical discrimination might be most effectively done by breaking down the in-group and out-group demarcation, forcing people to reconsider the personhood of those discriminated against. Depending on the values one holds and how strongly or loosely they are tied with in-group loyalty, different moral frames might be necessary to remind people of the discriminated applicants' personhood and moral standing and, thus, convince people to oppose statistical discrimination. To some, the most effective moral lens might be reminding people of the rights of persons—that people should not be harmed and that they should be treated as equals and with fairness. But for others, we might need to link such rights with in-group loyalty and, therefore, appeal to the (moral or even statistical) similarities between groups. Thus, to effectively convince people to oppose statistical discrimination, it is necessary to link the importance of personhood with the values people already endorse or hold. I will present the methods and results of this effort in the next chapter.

Chapter 6: Reducing the Acceptance of Statistical Discrimination

The previous chapters have shown the following: 1) People accept discrimination when and because it reflects accurate statistical group differences. 2) Acceptors justify ethnic recruitment discrimination based on perceived moral and cultural differences (negative stereotypes) that might negatively affect the workplace. 3) Many cases of statistical discrimination wrong the discriminatees, such as in the case of ethnic hiring discrimination. This brings me to the final research question (Q3): *How does morally reframing statistical discrimination influence people's acceptance hereof?*

As respondents indicate that moral and cultural differences between groups are the drivers of statistical group differences, I believe it is fruitful to stay in the moral lane when trying to reframe the case of statistical recruitment discrimination. However, since the results presented in Chapter 4 also show that people pass harsh moral judgments on employers who rely on inaccurate statistical group differences, statistical information in and of itself might alternatively effectively persuade people to decrease their acceptance of statistical recruitment discrimination.

This chapter builds on experimental methods and results, essentially testing two different mechanisms of reducing people's moral acceptance of ethnic recruitment discrimination. The first potential mechanism is that appealing to statistical differences influences people's acceptance by directly indicating which groups are efficient choices versus inefficient choices. This implies that statistical accuracy is at the core of the moral assessment because it is (perceived as) objective and efficient. In turn, positive statistical information about the discriminated would signal that the discriminatory action does not constitute an objective and efficient practice. The second potential mechanism is that statistical accuracy serves as a demarcation of groups, prompting people to assess who belongs to the (perceived) moral and cultural in-group. This implies that it will not be enough to counter negative statistical information about a group with positive statistical information as this still draws distinctions between groups and, therefore, makes them salient. Rather, this group loyalty-based assessment might be curbed by morally reframing the discriminated group either as people worth protecting against the harms of discrimination or, alternatively, as people who are part of a moral in-group. Both are ways to remind people of the moral standing of those discriminated against.

The rest of the chapter is structured as follows: First, I give a brief outline of moral reframing and informational framing. Second, I present the experimental method employed to test the two potential mechanisms of reducing people's acceptance of statistical discrimination. Third, I present the results, and fourth, I reflect on the implications of the results for the two hypothesized mechanisms and for the central argument of the dissertation.

6.1 Informational and Moral Frames

The two mechanisms that I hypothesized may drive the effect of appealing to accurate statistics rely on two different assumptions about people's opinion formation. The first mechanism presupposes that people are motivated by accuracy to form a "correct" or "optimal" judgment, whereas the second mechanism presupposes that people are motivated by existing opinions or values (Druckman, 2012). In the latter case, however, the existing beliefs need not be judgments about the specific policy or action, but rather existing beliefs in the moral values reflected by a policy (Bayes et al., 2020; Bayes & Druckman, 2021; Cusimano & Lombrozo, 2023). As a result, informational and statistical frames (serving to correct beliefs) have the potential to decrease acceptance of statistical discrimination if people have accuracy goals, and morally reframing statistical discrimination has the potential to decrease acceptance if people have directional value-based goals.

Many studies in the social sciences have investigated whether and when information can change people's minds (Druckman, 2022). Studies most closely related to opinions on ethnic recruitment discrimination are those that look into whether correcting voters' misperceptions about certain ethnic groups (or immigrants) can change their policy opinions, factual beliefs, or feelings toward the groups (Carnahan & Bergan, 2022; Grigorieff et al., 2020). However, the results are mixed as several studies find null effects, especially when looking at whether correcting misperceptions changes policy attitudes (Hopkins et al., 2019; Jørgensen & Osmundsen, 2022; Pedersen & Nielsen, 2022). This indicates that accuracy goals might not be the primary mode of opinion formation when making value judgments where ethnic groups are made salient.

Another strand of research explores the influence of moral values or moral rhetoric on opinion formation. Moralized political rhetoric might be perceived as a precarious tool because it has been found to produce a range of concerning effects, such as affective polarization (Simonsen & Bonikowski, 2022) and opposition to political compromises (Ryan 2017). Contrasting such results, the strategic use of moral values has also been shown to have no effects on polarization (Clifford and Simas 2022), to mobilize in-group partisans through

heightened positive emotions (Jung 2020), to increase perceptions of candidate sincerity (Clifford & Simas, 2022; Van Zant & Moore, 2015), and, most importantly to this project, to change people's policy opinions (see, e.g., Feinberg & Willer, 2019; Kalla et al., 2022; Kodapanakkal et al., 2022; Lau et al., 2021; Voelkel et al., 2023; Voelkel & Feinberg, 2018). An appeal to people's moral values thus holds promise for influencing people's opinions on political and social issues.

Recruitment discrimination is a salient and politicized social issue, which means that opinions on the matter might not be driven by a consequentialist calculus, i.e., trying to accurately estimate what the most optimal decision-making process would be. Instead, many people might factor in deeply held values. If people are directional rather than accuracy-oriented when it comes to ethnic recruitment discrimination, statistical information is not sufficient for changing people's judgments. This is because it implies that they should change their values (for instance, from loyalty to benefit). Instead, if moral appeals or frames stipulate that one's initial judgment (acceptance of discrimination) is unaligned with one's values, and one is motivated by directional (value) goals, then one would be motivated to change their judgment (rather than their value).

However, if people are driven by directional goals in their discrimination judgments, statistical information could bolster the effect of moral reframing (Simonsen & Bonikowski, 2022; Voelkel et al., 2022) because it gives evidence in line with what respondents are told should flow from their values. Thus, whereas statistical information might not be sufficient on its own to change people's minds, it might work after (or in tandem with) moral reframing because it provides another, albeit weaker, reason to change one's mind; i.e., to consistently endorse the value one is motivated by.

6.2 Method and Research Design

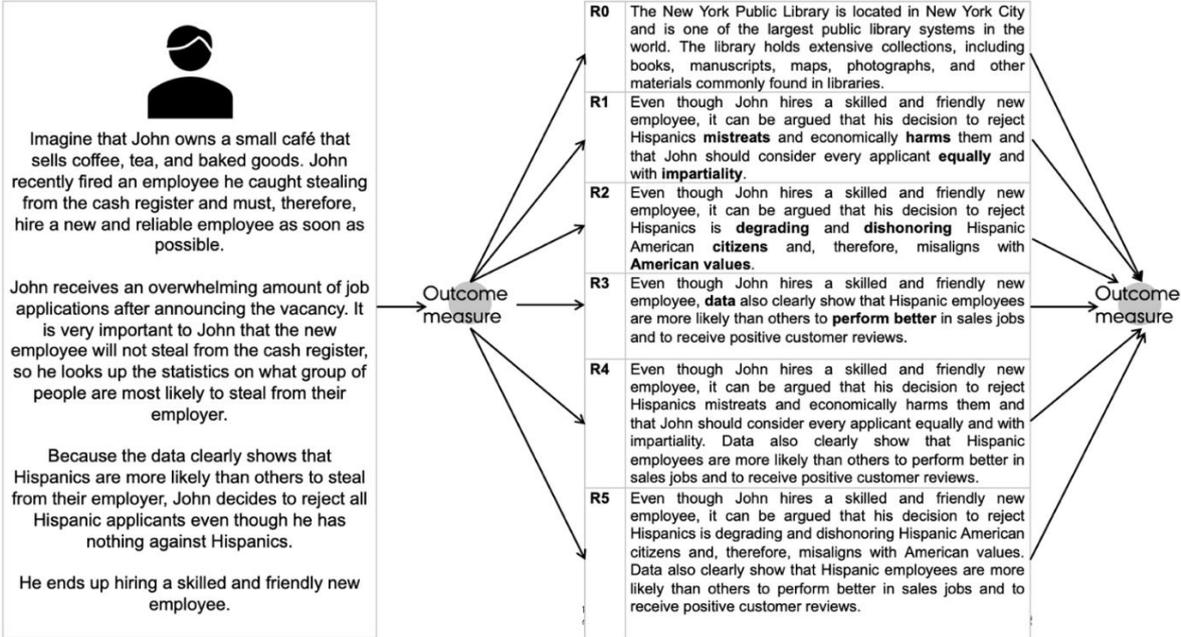
To test the effect of moral and statistical reframing, I conducted an online survey experiment with a between-subjects repeated measures design. I recruited an approximately representative sample (via CINT/Lucid) of 2,406 US adult respondents who were all presented with a case of statistical recruitment discrimination (rejecting Hispanics) and afterwards randomly assigned to one of five treatments. They encountered either a moral, statistical, or combination reframing of the recruitment decision. Whereas the statistical reframing condition would supposedly work similarly across the political spectrum (if people are generally accuracy-oriented), the moral reframing conditions are only likely to work if they appeal to moral values that people already endorse (Feinberg & Willer, 2019). The set of moral values people endorse generally

correlates strongly with political orientation (Graham et al., 2013; Kivikangas et al., 2021). For this reason, liberal values might not work effectively to persuade Republicans, and conservative values might not effectively persuade Democrats. The experiment therefore includes six different conditions:

1. A control condition with information about the New York Library.
2. A moral condition reframing the recruitment decision as violating liberal values of care and fairness.
3. A moral condition reframing the recruitment decision as violating conservative values of loyalty and sanctity.
4. A statistical condition informing respondents that Hispanics are better in sales and customer service.
5. A combination frame including the statistical and the liberal moral conditions.
6. A combination frame including the statistical and the conservative moral conditions.

I measured the dependent variable, moral acceptance, on a seven-point Likert scale twice: first after the initial presentation of the recruitment decision and second after the reframing conditions. Figure 5 depicts the flow of the experiment and the content of the reframing conditions.

Figure 5. Experimental Flow and Conditions



Note. Words in bold mark the moral buzzwords (liberal in R1 and conservative in R2) or the statistical/informational buzzwords (R3). The buzzwords were not in bold when presented to the respondents.

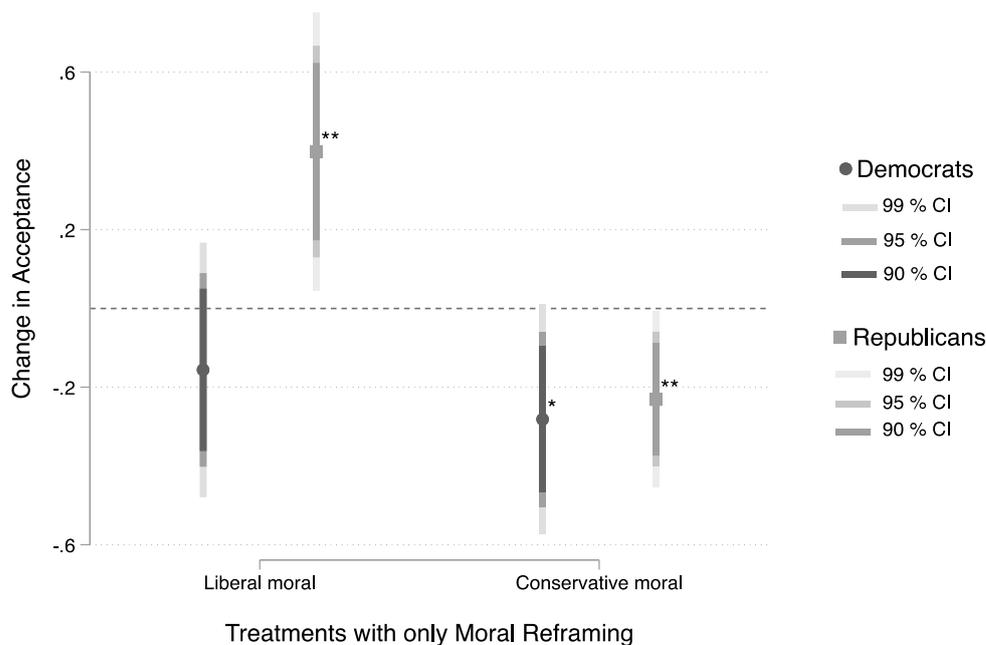
All demographic variables and potential moderator variables (described in Article C) were measured before the experimental part of the survey.

6.3. Results

To investigate whether moral or statistical reframing decreases people’s acceptance of discrimination, I used OLS regression to test whether the acceptance means across conditions were statistically significant from the acceptance mean of the control condition. I controlled for the pretreatment outcome measure of acceptance to increase statistical precision (Clifford et al., 2021).

I do not find that statistical information about Hispanics affects people’s acceptance in either direction ($p = 0.737$). Instead, I find that moral values influence people’s acceptance of the discriminatory recruitment decision. Still, only the conservative moral reframing conditions *decrease* acceptance across subgroups with different political orientations (for both conservative conditions: $p = 0.000$). Liberal moral values increase Republicans’ average acceptance of the discriminatory recruitment decision ($p = 0.004$) and do not change Democrats’ average acceptance in either direction ($p = 0.214$). Conservative values decrease Republicans’ and Democrats’ average acceptance ($p = 0.008$ and $p = 0.013$, respectively). These results are shown in Figure 6.

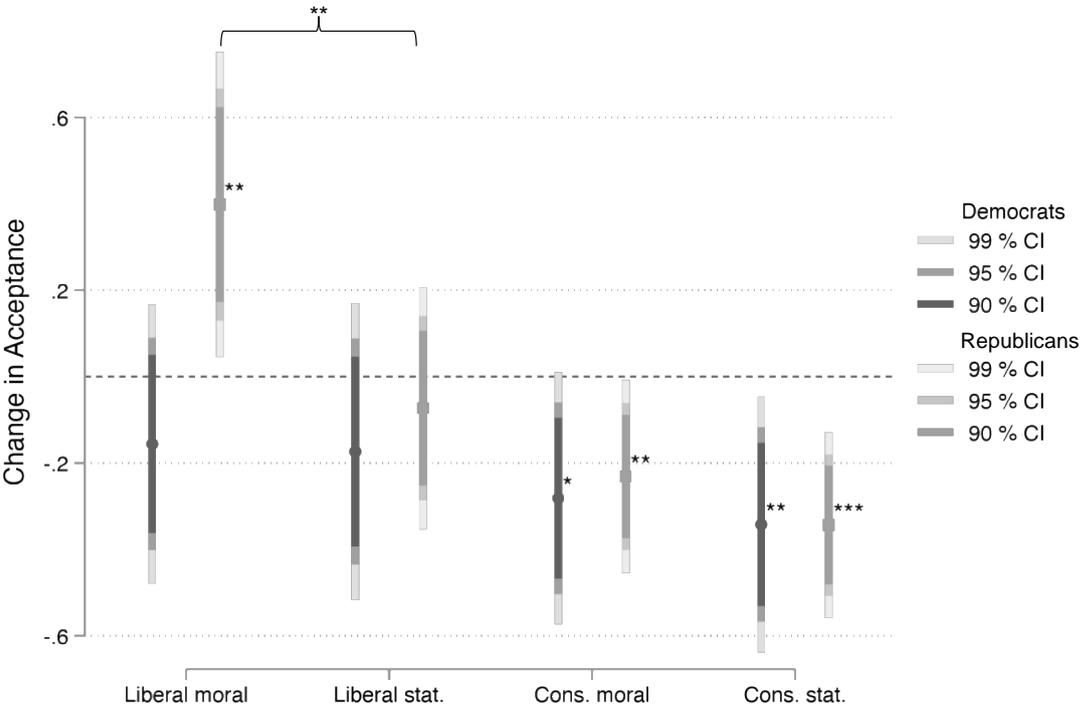
Figure 6. ATE of Moral Reframing across Political Identification



Note. Figure 6 shows the coefficient estimates for the acceptance outcome across the two moral reframing conditions. The control group (placebo) is the reference (the dotted o-line). Bars indicate 99%, 95%, and 90% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Democrats $n = 1,001$. Republicans $n = 912$.

These results suggest that people may be motivated by values when judging discrimination and are willing to change their minds *if* their former judgment is not as well-aligned with their values as initially assessed. Furthermore, whereas the effect size of combining the conservative moral frame with the statistical frame is larger than for the conservative moral frame alone, the difference between the two means is not statistically significant ($p = 0.211$). Interestingly, when combining the liberal moral frame with a statistical frame, it no longer increases Republicans' acceptance of discrimination. As such, whereas it might not backfire to add the statistical frame when morally reframing, it also does not significantly bolster the effect of the moral appeal. These results are shown in Figure 7.

Figure 7. ATE of Moral and Combination Frames across Political Identification



Note. Figure 7 shows the coefficient estimates for the acceptance outcome across the two framing conditions with liberal values and the two conditions with conservative values. The control group (placebo) is the reference (the dotted o-line). Bars indicate 99%, 95%, and 90% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Democrats $n = 1,001$. Republicans $n = 912$.

It is worth mentioning that among respondents identifying as true independents (leaning neither towards Democrats nor Republicans), the conservative moral frames did not produce any statistically significant effects ($p = 0.073$). However, the direction and magnitude of their assessment change were similar to those identified among Democrats and Republicans. The only statistically significant reframing condition among independents was the liberal

moral frame combined with the statistical frame ($p = 0.000$). This is somewhat puzzling, especially given that the liberal values did not persuade the target group most likely to subscribe to such values (Democrats). A speculative theory as to why I identify these effects among independents could be that the results were driven by a group understudied and poorly understood within the MFT framework and literature, namely libertarians. Libertarians have been shown to behave differently than endorsers of traditional liberal and conservative values, as they might be more pragmatic and freedom-oriented than what MFT has traditionally captured (Graham et al., 2013; Iyer et al., 2010). They are also the least inclined to hold strong group identities (ibid.). Whereas independents probably consist of a heterogeneous group (it may, for instance, also include self-identified socialists), the effect of the moral liberal and statistical combination frame could be driven by a substantial number of moral libertarians. This frame might very well resonate with them as it appeals to a pragmatic way of thinking (the fact that Hispanic citizens might be the most optimal choice given their performance statistics), *and* to the negative economic consequences of discrimination that limit the discriminatees' freedom. However, if such speculations are close to the truth, there remains the question of why neither the statistical frame nor the moral liberal frame is sufficient to influence independents' judgments.

To summarize, the results of the OLS regressions show that it is possible to decrease people's acceptance of statistical discrimination by reframing the discriminatory action as a violation of the conservative moral values of loyalty and sanctity. Strikingly, this effect holds up among Republicans *and* Democrats. The results further show that statistical frames are not sufficient on their own and might add very little bolstering to the effect of moral reframing. Nevertheless, it might be wise to include statistical frames when morally reframing statistical discrimination as it could cushion the backfiring or polarizing effects of moral reframing. Whereas I cannot once and for all conclude that people are motivated by directional value goals rather than accuracy goals when assessing discrimination, the results point toward directional goals as people are willing to change their judgments to fit their values, but not willing to change judgments based on statistical contra-evidence.

6.4 Reflections on Moral Frames and the Research Question

The results presented in this chapter enable me to answer the dissertation's third and final guiding research question: *How does morally reframing statistical discrimination influence people's acceptance hereof?* Morally reframing statistical recruitment discrimination as a violation of liberal values

polarizes across political orientation, and reframing it as a violation of conservative values decreases acceptance across political orientation. These findings also inform the dissertation's core research question by showing that information about statistical group differences cannot be used to decrease people's acceptance. Instead, we should look to moral values if we want to curb the acceptance of recruitment discrimination.

The results tie rather well with the findings presented in Chapter 4, identifying acceptors' justifications as centering around in-group loyalty and the perceived characteristics of the out-group. An approach that morally reframes statistical discrimination as a violation against loyal citizens who are not so different from members of the initially perceived in-group prompts people to identify an alternative moral interpretation of the discriminatory act. It serves as a reminder that those who are discriminated against are people who should be met with dignity and respect. As such, the loyalty framing invites people to include the discriminated applicants into their sphere of moral concern *and* the sanctity framing provides a reminder of the discriminatees' moral standing.

In Article C, I also discuss whether conservative values might persuade by virtue of being novel perspectives or arguments. Unlike liberal values focusing on equality and fairness, loyalty and sanctity are seldom connected with anti-discrimination argumentation, law, or rhetoric. As a result, respondents might have been pretreated and fatigued by liberal arguments but not by conservative arguments (see, e.g., Song & So, 2023). Nevertheless, values such as loyalty were prominent among acceptors of discrimination, so it seems unlikely that conservative frames would appear novel to this group.

Moral motivation (Curren & Ryan, 2020) can be an important antecedent of action and policy opinions. The results and discussions in Chapter 6 illustrate that moral values can effectively persuade people to decrease their acceptance of statistical discrimination. Moral arguments and values are thus important factors for understanding and explaining when and why people accept socially unwanted practices such as ethnic recruitment discrimination.

Chapter 7: Discussion and Conclusion

In the previous chapters, I have shown that Danes and Americans are more accepting of recruitment discrimination when the discriminated group is statistically different from the relevant comparator group (often the general population). I have also provided evidence for which moral values might underpin people's acceptance. Acceptors of discrimination ground their assessments in the values of efficiency and loyalty, especially focusing on the employer's right and duty to protect the workplace, and on the perceived (lacking) moral character of the discriminated group. Furthermore, I have provided an argument for why we should be hesitant to conclude that accurate statistical group differences make discrimination categorically more acceptable than other forms of discrimination. In short, I argue that cases of statistical discrimination often constitute what I term wrongful objectification. The latter is a grave form of objectification whereby a person is reduced to one or a few of their bodily or social properties and instrumentalized to reach a goal that is extraneous to the interaction-specific interests of that person. Since wrongful objectification provides a reason for why we should try to mitigate people's acceptance of statistics-based discrimination, I finally demonstrated that appealing to the conservative values of loyalty and sanctity is an effective way of doing so. In other words, I have demonstrated that it is possible to remind people of the costs of discrimination and thereby diminish the perceived value trade-off between harmful and unequal treatment on the one hand and in-group loyalty and efficiency on the other.

Taken all together, these results and arguments answer the research question of the dissertation. In the first chapter, I asked *what role statistical group differences play for the moral acceptability of discrimination*, and I provided a condensed answer in the form of a central argument. To reiterate and nuance that argument, I argue that statistical group differences play a significant role in people's assessments of discrimination, as they serve to make group identities salient and trigger emphasis on the values of loyalty, efficiency, and objectivity. I have, furthermore, argued that despite these empirical results, statistical group differences *should not* play a significant role in our moral assessments of discrimination. In other words, they do not alleviate the wrongs of targeting the victims of discrimination but might, on the contrary, add to the disrespect entailed in the action by wrongfully treating them as objects. Lastly, statistical group differences play a role in rationalizing acceptance of discrimination, but it is not clear that they work the other way around. Statistical

differences that reflect positively on the discriminated group do not seem to balance out statistical information that reflects negatively on the group. That might be because the existence of statistical differences triggers in-group loyalty as a value and any differences (positive or negative) solidify the notion that the groups *differ*. As a result, people still find a reason to use loyalty as an assessment value. Instead, the acceptance of statistical discrimination can be curbed by morally reframing discrimination, extending loyalty to the discriminated group and emphasizing the group's moral character and how that calls for a certain level of respect.

Whereas my dissertation sheds light on the role of statistical group differences for the acceptance of discrimination, it also invites further questions. Among those are questions about the underlying normative or maybe even paternalist motivation of the project, the limitations of the scope of the project, and, finally, about its real-world implications. I will devote the final part of this summary report to answering some of these questions.

7.1 Normativity, Empirics, and Paternalism

In my dissertation, I seek to understand and explain when and why people accept discrimination and also engage in a normative, critical assessment of such empirical findings. If people tend to accept discrimination based on accurate statistical group differences, it invites us to consider whether this is a problem, normatively speaking. In Chapter 5, I argue that it is morally problematic to accept discrimination just because the existence of statistical group differences allows recruiters to efficiently allocate their time and resources towards one social group at the expense of other groups. I therefore also seek to find a way to mitigate people's acceptance, which motivates the studies in Chapter 6. One might reasonably ask whether the normative elements of my project bias or distort my empirical investigations, and whether it is paternalistic to seek to mitigate people's acceptance of statistical discrimination (assuming that acting morally is in their best interest), now that I have found that their assessments are underpinned by moral values and considerations (and not, say, only psychological biases).

Regarding the balance between the normative and descriptive elements of the project, I am explicit about when I investigate an empirical and descriptive puzzle and when I take a normative approach. I have clarified why people's increased acceptance should not lead us to conclude that accurate statistical group differences alleviate the wrongs entailed in discriminatory practices. Moreover, the cases of recruitment discrimination experimentally analyzed are, in most instances, legally prohibited in the European Union and the

United States. In that sense, morally condemning these cases of discrimination is not legally or normatively controversial.

Additionally, my normative framework should be evaluated from what Lindauer (2020) terms the fruitfulness view (as mentioned in Chapter 2). One central evaluation parameter, according to Lindauer, concerns whether a framework provides a guideline for how to prevent unwanted behaviors (ibid: p. 2136). In other words, a normative concept can be empirically fruitful if, when internalized, it prevents unwanted behavior. My normative framework precisely seeks to identify and explain when and why statistical discrimination should be avoided. In the same vein, I provide an empirical example of how to practically influence people to internalize anti-discrimination norms, considering the current politicized nature of discrimination and the relative effect of different forms of appeal. The normative elements of my dissertation thus do not bias my empirical endeavors but rather qualify and assess the implications of my findings, and vice versa. The combination of normative and empirical insights improves the practical fruitfulness of the dissertation's central argument and enables me to thoroughly reflect on the implications of the empirical findings, thereby strengthening my dissertation's research project.

Next, some might question whether the latter part of the dissertation is problematically paternalist because I investigate how to convince people to oppose statistical discrimination. The experiment in Chapter 6 is, in essence, a persuasion study, as it seeks to influence people's opinions or mental states (Druckman, 2022). However, moral reframing importantly provides people with reasons, values, and thus an argument for why they should change their minds. To echo Fleisje (2023) and Tsai (2014), there is a normatively important difference between different modes of persuasion. Fleisje (2023) draws a distinction between persuading and convincing others, as the former often has paternalist tendencies while the latter does not. Similarly, Tsai (2014) argues that rational persuasion can be used both paternalistically and non-paternalistically. I would argue that my project is non-paternalistic (or at least not problematically paternalistic) because two crucial features are present: I am (1) respecting the receiver's ability to weigh the information, values, and arguments in a manner they see fit (ibid.; Shiffrin, 2000) and (2) seeking to influence people by appealing to normatively sound argumentation rather than triggering logical fallacies (Corner & Hahn, 2010). Whereas appealing to logical fallacies leverages people's psychological biases, appeals to normative arguments and values take people's capacity to evaluate and engage with arguments and information seriously. This does not mean that different moral frames do not prime different modes of thinking or different evaluation standards (such as whether respondents evaluate a practice by its harmful consequences or disrespect towards an in-group). Rather, different moral frames

and appeals prime different ways of thinking about an issue (Feinberg & Willer, 2019). Importantly, such differences are triggered by differences in the moral information provided rather than prompting logical fallacies (Corner & Hahn, 2010).

Moral reframing is therefore a rather non-deceptive persuasion technique as the moral values, and thus the suggested evaluation standard, are made explicit to its targets. I use an approach that offers respondents a moral lens or basis of evaluation but does not force them in any direction, as evident from the backfiring effects among Republicans. If people consistently believe that statistical discrimination is morally acceptable, they can indicate so.

7.2 Limitations and Future Research

In this section I share a few reflections on the limitations of my studies and how they invite future research to dive into statistics-based discrimination and moral values. First, it is important to discuss the dissertation's limitations in terms of external validity. I include data from Danish and American respondents, and, in both contexts, I investigate the acceptance of discrimination targeting a perceived immigrant group. Since immigration is strongly politicized and salient in Denmark and the United States, the political or ideological climate could very well be a contributing factor to people's willingness to discriminate against this specific group (Esses, 2021). It would be beneficial to test the effect of appealing to statistical group differences across contexts with different levels of immigration politicization and regarding different types of groups. It is likely that information about groups works differently depending on the pervasiveness of group stereotypes or whether information confirms or disconfirms stereotypes (see, e.g., Portmann, 2022).

Even if the results generalize across cultural contexts, I have traded mundane realism for internal validity (McDermott, 2011). While I try to get close to how people morally assess situations that occur in many individuals' lives, I do not measure everyday moral decision-making or behavior. Some studies have tried to investigate moral decision-making in the field, such as Kalla et al. (2022), whose field experiment found that encountering anti-abortion activism tailored to one's moral values and beliefs had an effect on people's inclination to engage in activism and potentially on policy attitudes. Similarly, field experiments aiming to inform citizens (or mitigate disinformation) have also found that people change their factual beliefs or behavior (Altmann et al., 2018; Aydin et al., 2018; Liebman & Luttmer, 2015; Larsen & Olsen, 2020). However, such field experiments rely on carefully thought-out interventions and do not necessarily get at day-to-day interactions between people. So, while it is reasonable to assume that people face information about statistical group

differences throughout their lives and that this might influence their acceptance of discriminatory behaviors or practices, we know very little about how and whether people moralize statistics-based discriminatory behavior when they encounter it in their daily lives.

Second, when investigating respondents' moral values and the effect of moral appeals on acceptance, it is not a given that conservative values of loyalty and sanctity would have the same causal effect as identified among American respondents. In the United States, conservatism and religion are more intimately linked than in Western Europe (Kivikangas et al., 2021). Furthermore, as identified among the Danish respondents' justifications, freedom of choice and employer rights/duties figure as central themes. In a Danish context, values connected to liberalism and libertarianism might better explain attitudes among many right-wingers than conservative values. Thus, in the future, it might be fruitful to investigate whether conservative values also decrease Western Europeans' acceptance of recruitment discrimination or whether alternative value appeals, such as liberty, better resonate with Western European citizens. As mentioned, Danes appear much more willing to accept recruitment discrimination than Americans and therefore it might be even more crucial to explore what it takes to decrease acceptance among Danes.

A third limitation and thus a potential avenue for future research concerns which mechanism best explains the effect of appealing to accurate statistical group differences. It is especially puzzling that group statistics can be used to justify acceptance of discrimination or increase people's acceptance while it does not seem to work the other way around. If people generally are motivated by their prior attitudes towards a group, statistics reflecting negatively on the targeted group would be used to rationalize discrimination by those with negative attitudes towards the group, while statistics reflecting positively on the targeted group would be used to condemn discrimination by those with positive attitudes towards the group. This type of motivated reasoning seems to conflict with the finding that there is no significant difference in how citizens with warm versus cold feelings towards the discriminated group react to the statistical information. Across the board, people update in the direction of accepting discrimination more, which might lead us to reconsider whether people could be driven by accuracy goals (see, e.g., Bisgaard, 2019; Hopkins et al., 2019; Jørgensen & Osmundsen, 2022; Schaeffer et al., 2023 for examples of when respondents update beliefs but not normative or political stances).

Another potential mechanism is tightly connected to the moral value of loyalty, namely ethnocentrism (Kinder & Kam, 2010). Even though my experiments do not manipulate the exposure to out-groups, they could make in-group and out-group dynamics more salient among those who get the

information about accurate statistical group differences. After all, this type of information not only conveys a descriptive fact about the world; it also (albeit latently) conveys a message that there are significant differences between groups that could be of moral significance. Furthermore, Newman et al. (2021) find that racial speech by political elites legitimizes or emboldens pre-existing racial prejudices among American citizens. Maybe statistical information works the same way? However, I do not find much support for this mechanism, as neither being part of the ethnic group discriminated against (or not) nor feelings towards the discriminated group moderate the effect of appealing to statistics. However, in line with Newman et al. (2021), it might be because the anti-discrimination norms are so deeply ingrained that we would not expect to see violations of such norms without a legitimization or moral license. Nevertheless, I do not find that the respondent's level of attention to social comparison moderates the effect of the statistical appeals (only the moral appeals). Future research into statistics-based discrimination could thus investigate what psychological mechanisms underpin the effect of appealing to statistical group differences.

Lastly, my normative framework of wrongful objectification faces some practical limitations. I have taken for granted that we often know what happens behind the scenes when recruitment or allocation decisions are made. In practice, however, people might rarely know the motives and processes behind discriminatory decisions. Most of the time, we are limited to scrutinizing the outcomes of, say, recruitment decisions as an indicator of whether a hiring process might have been morally suspicious. My framework might seem redundant if employers claim that they do not rely on statistical proxies. However, we generally face the difficult challenge of identifying when any discriminatory practice has occurred. Furthermore, I believe that accurate statistical group differences will often be used as a justification when we identify an action as discriminatory. As such, my normative contribution provides an argument for why statistical group differences rarely serve as an alleviating or justifying circumstance. Future normative work on discrimination would nevertheless benefit from critically engaging with the concept of statistical discrimination. Is it truly a distinct way of behaving with a distinct set of motivations, or should we move towards what I term statistics-based discrimination, where statistical group differences might enter very late in a discriminator's deliberation—much later than implicit and automatic biases and deeply held stereotypes?

7.3 Social and Political Implications

An upshot of my dissertation is that disseminating statistical information about societal groups could exacerbate discriminatory behaviors or attitudes in society. In this sense, facts are not just facts. They also indirectly function as normative triggers prompting people to evaluate recruitment discrimination based on the moral value of (in-group) loyalty. As such, when politicians use statistical facts about groups to support an argument or policy, they may also indirectly and inadvertently share the message that because people are statistically different, they are morally different or that some groups are less morally deserving than others. In turn, anti-discrimination organizations face a difficult challenge in curbing the (potentially increased) acceptance of discrimination. If they appeal to positively framed statistics about targeted groups, it is unlikely to mitigate people's acceptance. If they instead appeal to their egalitarian values, it might backfire and further increase people's acceptance of discrimination.

This means that political actors are left with two dilemmas. First, facts are important in politics, but without caution, they can trigger potentially worrying moral conclusions. Thus, it is important to ask when and how statistical facts about groups should be incorporated into political communication. Second, political communication that appeals to the moral values underpinning anti-discrimination beliefs could harm efforts to reduce discrimination. Instead, political actors could appeal to conservative values to effectively decrease people's acceptance of discrimination. Such actors may need to weigh communication authenticity against effectiveness.

English Summary

If the majority of us agree that discrimination is an unacceptable way to treat another person, why is it still a widespread social problem? While much research has examined what causes people to discriminate—often unintentionally—we know much less about when people accept discrimination. This dissertation therefore seeks answers to the latter. It does so by investigating the significance of statistical group differences for people's willingness to accept recruitment discrimination.

While other social science studies and theories often explain discriminatory dispositions as a result of psychological biases, organizational structures, or social dynamics, this dissertation points to an alternative thesis. Attitudes towards discrimination can instead be understood as an expression of moral values, and especially value trade-offs. Discrimination triggers a trade-off dilemma when it is based on statistical group differences. In such cases, the value of equal treatment is weighed against the values of efficiency and group loyalty.

The dissertation examines this proposition by first investigating when Americans and Danes accept recruitment discrimination. This is done using two experiments that test whether employer intent and information about statistical group differences influence people's acceptance of discrimination. The results demonstrate that information about statistical group differences increases the acceptance of discrimination. Additionally, the dissertation investigates which moral values Danes and Americans base their acceptance on. It turns out that values such as loyalty, freedom, objectivity, and efficiency are especially used to justify discrimination.

Next, a normative framework is presented to assess when statistical discrimination is morally unacceptable. The central principle of the framework concerns a specific form of disrespect—namely objectification—and demonstrates that statistical discrimination is morally unacceptable when it objectifies the discriminated party.

Finally, the dissertation examines what can be done to reduce people's acceptance of recruitment discrimination based on statistical group differences. This is done through an experiment that reframes discrimination as either being inconsistent with liberal values or conservative values, or by appealing to statistical group differences that present the discriminated group in a positive light. Surprisingly, conservative value appeals are effective in reducing the acceptance of statistical discrimination, while liberal values polarize opinions across party identification.

One implication of the dissertation's findings is that political actors face two major dilemmas. First, they must weigh whether the importance of communicating about statistical group differences outweighs the potential normative consequences it may have. Second, activists working to reduce discrimination must decide whether they want to appeal to values such as equality and justice, even though it may increase acceptance among the conservative segment of the population, or whether they instead wish to appeal to conservative values such as loyalty, thereby compromising authenticity.

Dansk Resumé

Hvis størstedelen af os er enige om, at diskrimination er en uacceptabel måde at behandle et andet menneske på, hvorfor er det så stadig et udbredt socialt problem? Mens meget forskning har undersøgt, hvad der får folk til at diskriminere—ofte uden intention herom—ved vi meget mindre om, hvornår folk accepterer diskrimination. Denne afhandling søger derfor svar på sidstnævnte. Det gør den ved at undersøge betydningen af statistiske gruppeforskelle for folks villighed til at acceptere ansættelsesdiskrimination.

Mens andre samfundsvidenskabelige studier og teorier ofte forklarer diskriminerende dispositioner som et resultat af psykologiske biases, organisatoriske strukturer eller sociale dynamikker, peger denne afhandling på en alternativ tese. Holdninger til diskrimination kan i stedet forstås som et udtryk for moralske værdier og især værditrade-offs. Diskrimination udløser et trade-off-dilemma, når det baserer sig på statistiske gruppeforskelle. I sådanne tilfælde sættes værdien af ligebehandling over for værdierne effektivitet og gruppeloyalitet.

Afhandlingen undersøger denne tese ved først at undersøge, hvornår amerikanere og danskere accepterer ansættelsesdiskrimination. Dette gøres ved hjælp af to eksperimenter, der tester, om arbejdsgiverintention og information om statistiske gruppeforskelle påvirker folks accept af diskrimination. Resultaterne demonstrerer, at information om statistiske gruppeforskelle øger accepten af diskrimination. Dertil undersøger afhandlingen også, hvilke moralske værdier danskere og amerikanere ligger til grund for deres accept. Det viser sig, at især værdierne loyalitet, frihed, objektivitet og effektivitet bruges til at retfærdiggøre diskrimination.

Dernæst opstilles et normativt framework til at vurdere, hvornår statistisk diskrimination er moralsk uacceptabel. Det centrale princip i frameworket omhandler en bestemt form for disrespekt—nemlig objektivisering—og demonstrerer, at statistisk diskrimination er moralsk uacceptabel, når det objektiviserer den diskriminerede part.

Slutteligt undersøger afhandlingen, hvad man kan gøre for at mindske folks accept af ansættelsesdiskrimination, der baserer sig på statistiske gruppeforskelle. Dette gøres via et eksperiment, som reframer diskrimination som enten i uoverensstemmelse med liberale værdier eller konservative værdier, eller ved at appellere til statistiske gruppeforskelle, som sætter den diskriminerede gruppe i et positivt lys. Overraskende nok er konservative værdiappeller effektive til at mindske accepten af statistisk diskrimination, mens liberale værdier polariserer holdninger på tværs af partiidentifikation.

En implikation af afhandlingens fund er, at politiske aktører står over for to store dilemmaer. For det første bør de afveje, hvornår vigtigheden af kommunikation om statistiske gruppeforskelle opvejer de potentielle normative konsekvenser, det kan have. For det andet bør aktivister, der arbejder for at mindske diskrimination, afveje, om de ønsker at appellere til værdier som lighed og retfærdighed, selvom det kan øge accepten blandt den konservative del af en befolkning, eller om de i stedet vil appellere til konservative værdier som loyalitet og dermed gå på kompromis med autenticitet.

Bibliography

- Al Ramiah, A., Hewstone, M., Dovidio, J. F., & Penner, L. A. (2010). The Social Psychology of Discrimination: Theory, Measurement and Consequences. In L. Bond, F. McGinnity, & H. Russell (Eds.), *Making equality count: Irish and international research measuring equality and discrimination*. Liffey.
- Albertsen, A., Hallsson, B. G., Lippert-Rasmussen, K., & Pedersen, V. M. L. (2023). Does harm or disrespect make discrimination wrong? An experimental approach. *Philosophical Psychology*, 1–26.
<https://doi.org/10.1080/09515089.2023.2289451>
- Alexander, L. (1992). What makes wrongful discrimination wrong? Biases, preferences, stereotypes, and proxies. *University of Pennsylvania Law Review*, 141(1), 149–219.
- Alexander, L. (2016). Is Wrongful Discrimination Really Wrong? *SSRN Electronic Journal*. <https://dx.doi.org/10.2139/ssrn.2909277>
- Altmann, S., Falk, A., Jäger, S., & Zimmermann, F. (2018). Learning about job search: A field experiment with job seekers in Germany. *Journal of Public Economics*, 164, 33–49. <https://doi.org/10.1016/j.jpubeco.2018.05.003>
- Altonji, J. G., & Pierret, C. R. (2001). Employer Learning and Statistical Discrimination. *The Quarterly Journal of Economics*, 116(1), 313–350.
<https://doi.org/10.1162/003355301556329>
- Andow, J. (2016). Reliable but not home free? What framing effects mean for moral intuitions. *Philosophical Psychology*, 29(6), 904–911.
<https://doi.org/10.1080/09515089.2016.1168794>
- Andrighetto, L., Baldissarri, C., & Volpato, C. (2017). (Still) Modern Times: Objectification at work. *European Journal of Social Psychology*, 47(1), 25–35.
<https://doi.org/10.1002/ejsp.2190>
- Angrist, J. D., & Pischke, J.-S. (2009). Chapter 2. The Experimental Ideal. In *Mostly Harmless Econometrics* (pp. 11–24). Princeton University Press.
<https://doi.org/10.1515/9781400829828-007>
- Arceneaux, K. (2010). The Benefits of Experimental Methods for the Study of Campaign Effects. *Political Communication*, 27(2), 199–215.
<https://doi.org/10.1080/10584601003709407>
- Arneson, R. (2017). Discrimination and Harm. In K. Lippert-Rasmussen (Ed.), *The Routledge Handbook of the Ethics of Discrimination* (1st ed., pp. 151–163). Routledge. <https://doi.org/10.4324/9781315681634-15>
- Arneson, R. J. (2006). What Is Wrongful Discrimination? *San Diego Law Review*, 43(4), 775–808.
- Arrow, K. J. (1973). The Theory of Discrimination. In O. Ashenfelter & A. Rees, *Discrimination in Labour Markets*. Princeton University Press.
- Auspurg, K., Schneck, A., & Hinz, T. (2019). Closed doors everywhere? A meta-analysis of field experiments on ethnic discrimination in rental housing

- markets. *Journal of Ethnic and Migration Studies*, 45(1), 95–114.
<https://doi.org/10.1080/1369183X.2018.1489223>
- Aydin, E., Brounen, D., & Kok, N. (2018). Information provision and energy consumption: Evidence from a field experiment. *Energy Economics*, 71, 403–410.
<https://doi.org/10.1016/j.eneco.2018.03.008>
- Barak-Corren, N., Tsay, C.-J., Cushman, F., & Bazerman, M. H. (2018). If You're Going to Do Wrong, At Least Do It Right: Considering Two Moral Dilemmas at the Same Time Promotes Moral Consistency. *Management Science*, 64(4), 1528–1540. <https://doi.org/10.1287/mnsc.2016.2659>
- Bayes, R., & Druckman, J. N. (2021). Motivated reasoning and climate change. *Current Opinion in Behavioral Sciences*, 42, 27–35.
<https://doi.org/10.1016/j.cobeha.2021.02.009>
- Bayes, R., Druckman, J. N., Goods, A., & Molden, D. C. (2020). When and How Different Motives Can Drive Motivated Political Reasoning. *Political Psychology*, 41(5), 1031–1052. <https://doi.org/10.1111/pops.12663>
- Beatty, T. K., & Sommers, D. E. (2008). Discrimination in Europe. Evidence from the Rental Market. *Discussion Papers*, 547.
- Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.
- Beeghly, E. (2018). Failing to Treat Persons as Individuals. *Ergo*, 5(26).
<https://doi.org/10.3998/ergo.12405314.0005.026>
- Bell, E. (2021). Deserving to Whom? Investigating Heterogeneity in the Impact of Social Constructions of Target Populations on Support for Affirmative Action. *Policy Studies Journal*, 49(1), 268–299. <https://doi.org/10.1111/psj.12347>
- Bengson, J. (2013). Experimental Attacks on Intuitions and Answers. *LXXXVI*(3), 39.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774.
<https://doi.org/10.21105/joss.00774>
- Berndt Rasmussen, K. (2019). Harm and Discrimination. *Ethical Theory and Moral Practice*, 22(4), 873–891. <https://doi.org/10.1007/s10677-018-9908-4>
- Bertrand, M., & Duflo, E. (2017). Field Experiments on Discrimination. In *Handbook of Economic Field Experiments* (Vol. 1, pp. 309–393). Elsevier.
<https://doi.org/10.1016/bs.hefe.2016.08.004>
- Bisgaard, M. (2019). How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning. *American Journal of Political Science*, 63(4), 824–839.
<https://doi.org/10.1111/ajps.12432>
- Bohren, J. A., Haggag, K., Imas, A., & Pope, D. G. (2023). Inaccurate Statistical Discrimination: An Identification Problem. *Review of Economics and Statistics*, 1–45. https://doi.org/10.1162/rest_a_01367
- Borinca, I., Van Assche, J., Gronfeldt, B., Sainz, M., Anderson, J., & Taşbaşı, E. H. O. (2023). Dehumanization of outgroup members and cross-group

- interactions. *Current Opinion in Behavioral Sciences*, 50, 101247.
<https://doi.org/10.1016/j.cobeha.2023.101247>
- Brodeur, A., Cook, N., & Heyes, A. (2020). Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics. *American Economic Review*, 110(11), 3634–3660. <https://doi.org/10.1257/aer.20190687>
- Bullock, J. G. (2011). Elite Influence on Public Opinion in an Informed Electorate. *American Political Science Review*, 105(3), 496–515.
<https://doi.org/10.1017/S0003055411000165>
- Bunel, M., & Tovar, E. (2021). Profit, morality and discrimination. *APPLIED ECONOMICS*, 53(40), 4692–4712.
- Carnahan, D., & Bergan, D. E. (2022). Correcting the Misinformed: The Effectiveness of Fact-checking Messages in Changing False Beliefs. *POLITICAL COMMUNICATION*, 39(2), 166–183.
- Cath, Y. (2016). Reflective Equilibrium. In H. Cappelen, T. S. Gendler, & J. Hawthorne (Eds.), *The Oxford Handbook of Philosophical Methodology* (1st ed., pp. 213–230). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668779.013.32>
- Christensen, G. S., Freese, J., & Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science*. University of California Press.
- Clifford, S., Sheagley, G., & Piston, S. (2021). Increasing Precision without Altering Treatment Effects: Repeated Measures Designs in Survey Experiments. *American Political Science Review*, 115(3), 1048–1065.
<https://doi.org/10.1017/S0003055421000241>
- Clifford, S., & Simas, E. N. (2022). Moral Rhetoric, Extreme Positions, and Perceptions of Candidate Sincerity. *Political Behavior*.
<https://doi.org/10.1007/s11109-022-09835-w>
- Corner, A., & Hahn, U. (2010). Message Framing, Normative Advocacy and Persuasive Success. *Argumentation*, 24(2), 153–163. <https://doi.org/10.1007/s10503-009-9159-6>
- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66, 93–99. <https://doi.org/10.1016/j.jesp.2015.10.002>
- Curren, R., & Ryan, R. M. (2020). Moral self-determination: The nature, existence, and formation of moral motivation. *Journal of Moral Education*, 49(3), 295–315. <https://doi.org/10.1080/03057240.2020.1793744>
- Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology*, 6, 97–103. <https://doi.org/10.1016/j.copsy.2015.06.003>
- Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm. *Psychological Science*, 17(12), 1082–1089. <https://doi.org/10.1111/j.1467-9280.2006.01834.x>
- Cusimano, C., & Lombrozo, T. (2023). People recognize and condone their own morally motivated reasoning. *Cognition*, 234, 105379.
<https://doi.org/10.1016/j.cognition.2023.105379>

- Da Silva, M. (2021). The Traces Left Behind: On Appropriate Responses to Right Acts with Wrong Features. *Social Theory and Practice*, 47(1), 63–89. <https://doi.org/10.5840/soctheorpract2021115112>
- Dafoe, A., Zhang, B., & Caughey, D. (2018). Information Equivalence in Survey Experiments. *Political Analysis*, 26, 399–416.
- Daniels, N. (1979). Wide Reflective Equilibrium and Theory Acceptance in Ethics. *The Journal of Philosophy*, 76(5), 256. <https://doi.org/10.2307/2025881>
- DeBell, M. (2017). Polarized Opinions on Racial Progress and Inequality: Measurement and Application to Affirmative Action Preferences. *Political Psychology*, 38(3), 481–498. <https://doi.org/10.1111/pops.12342>
- Dreber, A., & Johannesson, M. (2019). Statistical Significance and the Replication Crisis in the Social Sciences. In A. Dreber & M. Johannesson, *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190625979.013.461>
- Druckman, J. N. (2012). THE POLITICS OF MOTIVATION. *Critical Review*, 24(2), 199–216. <https://doi.org/10.1080/08913811.2012.711022>
- Druckman, J. N. (2022). A Framework for the Study of Persuasion. *Annual Review of Political Science*, 25(1), 65–88. <https://doi.org/10.1146/annurev-polisci-051120-110428>
- Druckman, J. N., Green, D. P., Kuklinski, J. H., & Lupia, A. (2011). Experiments: An Introduction to Core Concepts. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (1st ed., pp. 15–26). Cambridge University Press. <https://doi.org/10.1017/CBO9780511921452.002>
- Eidelson, B. (2013). Treating People as Individuals. In D. Hellman & S. Moreau (Eds.), *Philosophical Foundations of Discrimination Law* (pp. 203–227). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199664313.003.0011>
- Eidelson, B. (2015). *Discrimination and disrespect* (First edition). Oxford University Press.
- Esses, V. M. (2021). Prejudice and Discrimination Toward Immigrants. *Annual Review of Psychology*, 72(1), 503–531. <https://doi.org/10.1146/annurev-psych-080520-102803>
- Evans, A. M., & Krueger, J. I. (2009). The Psychology (and Economics) of Trust. *Social and Personality Psychology Compass*, 3(6), 1003–1017. <https://doi.org/10.1111/j.1751-9004.2009.00232.x>
- Feinberg, M., & Willer, R. (2019). Moral reframing: A technique for effective and persuasive communication across political divides. *Social and Personality Psychology Compass*, 13(12). <https://doi.org/10.1111/spc3.12501>
- Fibbi, R., Midtbøen, A. H., & Simon, P. (2021). Theories of Discrimination. In R. Fibbi, A. H. Midtbøen, & P. Simon, *Migration and Discrimination* (pp. 21–41). Springer International Publishing. https://doi.org/10.1007/978-3-030-67281-2_3

- Fleisje, A. (2023). Paternalistic persuasion: Are doctors paternalistic when persuading patients, and how does persuasion differ from convincing and recommending? *Medicine, Health Care and Philosophy*, 26(2), 257–269. <https://doi.org/10.1007/s11019-023-10142-2>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505. <https://doi.org/10.1126/science.1255484>
- Gerber, A., & Malhotra, N. (2008). Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals. *Quarterly Journal of Political Science*, 3(3), 313–326. <https://doi.org/10.1561/100.00008024>
- Gervais, S. J. (Ed.). (2013). *Objectification and (De)Humanization* (Vol. 60). Springer New York. <https://doi.org/10.1007/978-1-4614-6959-9>
- Gilens, M. (2001). Political Ignorance and Collective Policy Preferences. *American Political Science Review*, 95(2), 379–396. <https://doi.org/10.1017/S0003055401002222>
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory. In *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Elsevier. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- Gray, K., MacCormack, J. K., Henry, T., Banks, E., Schein, C., Armstrong-Carter, E., Abrams, S., & Muscatell, K. A. (2022). The affective harm account (AHA) of moral judgment: Reconciling cognition and affect, dyadic morality and disgust, harm and purity. *Journal of Personality and Social Psychology*, 123(6), 1199–1222. <https://doi.org/10.1037/pspa0000310>
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364–371. <https://doi.org/10.1016/j.cognition.2009.02.001>
- Grigorieff, A., Roth, C., & Ubfal, D. (2020). Does Information Change Attitudes Toward Immigrants? *Demography*, 57(3), 1117–1143. <https://doi.org/10.1007/s13524-020-00882-8>
- Halldenius, L. (2018). Discrimination and Irrelevance. In K. Lippert-Rasmussen (Ed.), *The Routledge Handbook of the Ethics of Discrimination* (1st ed., pp. 108–118). Routledge. <https://doi.org/10.4324/9781315681634-11>
- Hellman, D. (2008). Demeaning and Wrongful Discrimination. In *When is discrimination wrong?* (pp. 34–58). Harvard University Press.
- Hopkins, D. J., Sides, J., & Citrin, J. (2019). The Muted Consequences of Correct Information about Immigration. *The Journal of Politics*, 81(1), 315–320. <https://doi.org/10.1086/699914>
- Hudson, R. (2023). Explicating Exact versus Conceptual Replication. *Erkenntnis*, 88(6), 2493–2514. <https://doi.org/10.1007/s10670-021-00464-z>

- Iyer, R., Koleva, S., Graham, J., Ditto, P. H., & Haidt, J. (2010). Understanding Libertarian Morality: The Psychological Roots of an Individualist Ideology. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1665934>
- Jørgensen, F. J., & Osmundsen, M. (2022). Correcting Citizens' Misperceptions about non-Western Immigrants: Corrective Information, Interpretations, and Policy Opinions. *Journal of Experimental Political Science*, 9(1), 64–73. <https://doi.org/10.1017/XPS.2020.35>
- Jung, J. (2020). The Mobilizing Effect of Parties' Moral Rhetoric. *American Journal of Political Science*, 64(2), 341–355. <https://doi.org/10.1111/ajps.12476>
- Kalla, J. L., Levine, A. S., & Broockman, D. E. (2022). Personalizing Moral Reframing in Interpersonal Conversation: A Field Experiment. *The Journal of Politics*, 84(2), 1239–1243. <https://doi.org/10.1086/716944>
- Khan, S. R., & Lambert, A. J. (2001). Perceptions of Rational Discrimination: When Do People Attempt to Justify Race-Based Prejudice? *Basic and Applied Social Psychology*, 23(1), 42–53. https://doi.org/10.1207/S15324834BASP2301_4
- Kinder, D. R., & Kam, C. D. (2010). *Us Against Them: Ethnocentric Foundations of American Opinion*. University of Chicago Press. ProQuest Ebook Central. <https://ebookcentral.proquest.com/lib/asb/detail.action?docID=515719>
- Kivikangas, J. M., Fernández-Castilla, B., Järvelä, S., Ravaja, N., & Lönnqvist, J.-E. (2021). Moral foundations and political orientation: Systematic review and meta-analysis. *Psychological Bulletin*, 147(1), 55–94. <https://doi.org/10.1037/bul0000308>
- Knight, C. (2017). Reflective Equilibrium. In A. Blau (Ed.), *Methods in Analytical Political Theory* (1st ed., pp. 46–64). Cambridge University Press. <https://doi.org/10.1017/9781316162576.005>
- Kodapanakkal, R. I., Brandt, M. J., Kogler, C., & Van Beest, I. (2022). Moral Frames Are Persuasive and Moralize Attitudes; Nonmoral Frames Are Persuasive and De-Moralize Attitudes. *Psychological Science*, 33(3), 433–449. <https://doi.org/10.1177/09567976211040803>
- Kronfeldner, M. (2021). INTRODUCTION Mapping dehumanization studies. In *THE ROUTLEDGE HANDBOOK OF DEHUMANIZATION* (Routledge).
- Kumar, V., & May, J. (2018). How to Debunk Moral Beliefs. In *Methodology and Moral Philosophy* (1st ed.). Routledge. <https://www.taylorfrancis.com/books/9780429839245>
- Larsen, M. V. (2021). How Do Voters Hold Politicians Accountable for Personal Welfare? Evidence of a Self-Serving Bias. *The Journal of Politics*, 83(2), 740–752. <https://doi.org/10.1086/710325>
- Lau, J. D., Song, A. M., Morrison, T., Fabinyi, M., Brown, K., Blythe, J., Allison, E. H., & Adger, W. N. (2021). Morals and climate decision-making: Insights from social and behavioural sciences. *Current Opinion in Environmental Sustainability*, 52, 27–35. <https://doi.org/10.1016/j.cosust.2021.06.005>
- Levy, J. S. (2008). Case Studies: Types, Designs, and Logics of Inference. *Conflict Management and Peace Science*, 25(1), 1–18. <https://doi.org/10.1080/07388940701860318>

- Liebman, J. B., & Luttmer, E. F. P. (2015). Would People Behave Differently If They Better Understood Social Security? Evidence from a Field Experiment. *American Economic Journal: Economic Policy*, 7(1), 275–299. <https://doi.org/10.1257/pol.20120081>
- Lindauer, M. (2020). Experimental philosophy and the fruitfulness of normative concepts. *Philosophical Studies*, 177(8), 2129–2152. <https://doi.org/10.1007/s11098-019-01302-3>
- Lippert-Rasmussen, K. (2011). “We are all Different”: Statistical Discrimination and the Right to be Treated as an Individual. *The Journal of Ethics*, 15(1–2), 47–59. <https://doi.org/10.1007/s10892-010-9095-6>
- Lippert-Rasmussen, K. (2013). *Born Free and Equal?: A Philosophical Inquiry into the Nature of Discrimination* (Published to Oxford Scholarship Online: January 2014). Cary: Oxford University Press. <https://oxford-universitypress-scholarship-com.ez.statsbiblioteket.dk:12048/view/10.1093/acprof:oso/9780199796113.001.0001/acprof-9780199796113>
- Lippert-Rasmussen, K. (2018). *THE ROUTLEDGE HANDBOOK OF THE ETHICS OF DISCRIMINATION*. Routledge/Taylor & Francis Group.
- Maitzen, S. (1991). The Ethics of Statistical Discrimination: *Social Theory and Practice*, 17(1), 23–45. <https://doi.org/10.5840/soctheorpract199117114>
- McDermott, D. (2008). Analytical political philosophy. In D. Leopold & M. Stears (Eds.), *Political Theory: Methods and Approaches* (pp. 11–28). Oxford University Press. <https://ebookcentral-proquest-com.ez.statsbiblioteket.dk/lib/asb/detail.action?docID=415667>
- McDermott, R. (2011). Internal and External Validity. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (1st ed., pp. 27–40). Cambridge University Press. <https://doi.org/10.1017/CBO9780511921452.003>
- Nagel, J. (2013). Defending the Evidential Value of Epistemic Intuitions: A Reply to Stich. *Philosophy and Phenomenological Research*, 87(1), 179–199. <https://doi.org/10.1111/phpr.12008>
- Newman, B., Merolla, J. L., Shah, S., Lemi, D. C., Collingwood, L., & Ramakrishnan, S. K. (2021). The Trump Effect: An Experimental Investigation of the Emboldening Effect of Racially Inflammatory Elite Communication. *British Journal of Political Science*, 51(3), 1138–1159. <https://doi.org/10.1017/S0007123419000590>
- Nussbaum, M. (1995). Objectification. *Philosophy & Public Affairs*, 24(4), 249–291.
- Paulo, N. (2020). The Unreliable Intuitions Objection Against Reflective Equilibrium. *The Journal of Ethics*, 24(3), 333–353. <https://doi.org/10.1007/s10892-020-09322-6>
- Pedersen, M. J., & Nielsen, V. L. (2022). Understanding Discrimination: Outcome-Relevant Information Does Not Mitigate Discrimination. *Social Problems*, spac006. <https://doi.org/10.1093/socpro/spac006>

- Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review*, 62(4), 659–661.
- Portmann, L. (2022). Do Stereotypes Explain Discrimination Against Minority Candidates or Discrimination in Favor of Majority Candidates? *British Journal of Political Science*, 52(2), 501–519.
<https://doi.org/10.1017/S0007123420000800>
- Rawls, J. (1999). *A theory of justice* (Rev. ed.). Belknap Press of Harvard University Press.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Roth, J., Deutsch, R., & Sherman, J. W. (2019). Automatic Antecedents of Discrimination. *European Psychologist*, 24(3), 219–230. <https://doi.org/10.1027/1016-9040/a000321>
- Ryan, T. J. (2017). No Compromise: Political Consequences of Moralized Attitudes. *American Journal of Political Science*, 61(2), 409–423.
<https://doi.org/10.1111/ajps.12248>
- Sartori, G. (1970). Concept Misformation in Comparative Politics. *American Political Science Review*, 64(4), 1033–1053. <https://doi.org/10.2307/1958356>
- Schaeffer, M., Krakowski, K., & Olsen, A. L. (2023). *Correcting Misperceptions about Ethno-Racial Discrimination: The Limits of Evidence-Based Awareness Raising to Promote Support for Equal-Treatment Policies*.
<https://doi.org/10.31235/osf.io/be9pf>
- Schauer, F. (2006). *Profiles, Probabilities, and Stereotypes*. Harvard University Press.
- Schauer, F. (2018). Statistical (And Non-Statistical) Discrimination. In K. Lippert-Rasmussen (Ed.), *The Routledge Handbook of the Ethics of Discrimination* (1st ed., pp. 42–53). Routledge. <https://doi.org/10.4324/9781315681634-5>
- Shiffrin, S. V. (2000). Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy & Public Affairs*, 29(3), 205–250.
<https://doi.org/10.1111/j.1088-4963.2000.00205.x>
- Sides, J. (2016). Stories or Science? Facts, Frames, and Policy Attitudes. *American Politics Research*, 44(3), 387–414. <https://doi.org/10.1177/1532673X15610190>
- Simonsen, K. B., & Bonikowski, B. (2022). Moralizing Immigration: Political Framing, Moral Conviction, and Polarization in the United States and Denmark. *Comparative Political Studies*, 55(8), 1403–1436.
<https://doi.org/10.1177/00104140211060284>
- Simonsen, K. B., & Widmann, T. (2023). *When Do Political Parties Moralize? A Cross-National Study of the Strategic Use of Moral Language in Political Communication on Immigration* [Preprint]. Open Science Framework.
<https://doi.org/10.31219/osf.io/hxd5a>
- Sommer Hansen, R., & Midtgaard, S. F. (Eds.). (2016). *Metode i normativ politisk teori*. Samfundslitteratur.

- Song, H., & So, J. (2023). Message fatigue beyond the health message context: A replication and further extension of So et al. (2017). *Human Communication Research*, 49(3), 339–344. <https://doi.org/10.1093/hcr/hqado21>
- Stefan, M., Holzmeister, F., Müllauer, A., & Kirchler, M. (2018). Ethnical discrimination in Europe: Field evidence from the finance industry. *PloS One*, 13(1).
- Stich, S., & Tobia, K. P. (2016). Experimental Philosophy and the Philosophical Tradition. In J. Sytsma & W. Buckwalter (Eds.), *A Companion to Experimental Philosophy* (pp. 3–21). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118661666.ch1>
- Tilcsik, A. (2021). Statistical Discrimination and the Rationalization of Stereotypes. *American Sociological Review*, 86(1), 93–122.
- Title VII of the Civil Rights Act of 1964. (n.d.). *Title VII of the Civil Rights Act of 1964*. 42 U.S.C. § 2000e et seq.
- Tsai, G. (2014). Rational Persuasion as Paternalism. *Philosophy & Public Affairs*, 42(1), 78–112. <https://doi.org/10.1111/papa.12026>
- Vaes, J., Leyens, J.-P., Paola Paladino, M., & Pires Miranda, M. (2012). We are human, they are not: Driving forces behind outgroup dehumanisation and the humanisation of the ingroup. *European Review of Social Psychology*, 23(1), 64–106. <https://doi.org/10.1080/10463283.2012.665250>
- Van Zant, A. B., & Moore, D. A. (2015). Leaders' Use of Moral Justifications Increases Policy Support. *Psychological Science*, 26(6), 934–943. <https://doi.org/10.1177/0956797615572909>
- Vinæs Larsen, M., & Leth Olsen, A. (2020). Reducing Bias in Citizens' Perception of Crime Rates: Evidence from a Field Experiment on Burglary Prevalence. *The Journal of Politics*, 82(2), 747–752. <https://doi.org/10.1086/706595>
- Voelkel, J. G., & Feinberg, M. (2018). Morally Reframed Arguments Can Affect Support for Political Candidates. *Social Psychological and Personality Science*, 9(8), 917–924. <https://doi.org/10.1177/1948550617729408>
- Voelkel, J. G., Malik, M., Redekopp, C., & Willer, R. (2022). Changing Americans' Attitudes about Immigration: Using Moral Framing to Bolster Factual Arguments. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 73–85. <https://doi.org/10.1177/00027162221083877>
- Voelkel, J. G., Mernyk, J. S., & Willer, R. (2023). Moral reframing increases support for economically progressive candidates. *PNAS Nexus*, 2(6), pgad154. <https://doi.org/10.1093/pnasnexus/pgad154>
- Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, 120(2), 202–214. <https://doi.org/10.1016/j.cognition.2011.04.005>
- Zschornt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7), 1115–1134.