Pretty Biased?

How Physical Appearance Shapes Patterns of Ethno-Religious Discrimination

Joshua Hellyer

Inaugural dissertation submitted in partial fulfillment
of the requirements for the degree
Doctor of Social Sciences
at the University of Mannheim

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *Don't judge a book by its cover.*
> *Beauty is only skin deep.*
> *It's what's inside that counts.*
> *Clothes don't make the man.*
> *Looks can be deceiving.*

We know from these and other common idioms that we should not judge others solely by their appearance. Yet, these cultural norms are necessary precisely because we humans are quick to make assumptions about others based on how they look (Todorov, 2017; Zebrowitz & Montepare, 2015). After only 100 milliseconds of looking at a person's face, we form lasting first impressions of their character traits that may influence how we treat that person moving forward (Willis & Todorov, 2006). Even though these first impressions are not always correct, the accessibility of facial cues makes them especially powerful in impression formation (Jaeger et al., 2019; Olivola & Todorov, 2010). In contemporary diverse societies where we frequently interact with people we do not know, we rely on facial and bodily cues as a heuristic to guide our interactions.

Many of the judgments we make from others' appearance are relatively benign. When addressing a stranger, one might use spontaneous appraisals of the person's gender, age, and status to determine how to politely address them. However, these assessments can also be prejudiced, leading to differential treatment and possibly discrimination against people belonging to visible minority groups. A Berlin shopkeeper might assume that a non-white customer cannot speak German, opting to speak to them in English instead (Williams, 2024). One might avoid coming into contact with other people based on a judgment of their race (Dietrich & Sands, 2023; Zhang et al., 2022), or decide not to help a hijab-wearing stranger due to perceptions about her religion (Choi et al., 2019). While the immediate consequences of any one of these situations might be minor, exposure to many such "microaggressions" over time has been linked to lower well-being for members of visible minority groups (Costa et al., 2023).

Appearance-based judgments also underlie more immediately consequential decisions. We might use someone's physical appearance to make assumptions about less visible qualities, such as their abilities and character traits. For example, a large

body of literature suggests that attractive people receive a variety of benefits on the labor market, suggesting that employers view them as more productive (Nault et al., 2020), more competent (Kühn & Wolbring, 2024), and more intelligent (L. A. Jackson et al., 1995). Similarly, voters also seem to prefer more attractive candidates for political office (Jäckle & Metz, 2017; Rosar et al., 2008). We may also find more attractive people to be less threatening: several studies show that more attractive people are less likely to be found guilty in court (Mazzella & Feingold, 1994), and receive more lenient sentencing when they are charged with a crime (Beaver et al., 2019; B. D. Johnson & King, 2017).

As evidenced by these diverse examples, appearance is a powerful cue, but also an extremely broad one. When we look at a person, we automatically assign them to a variety of social categories, especially (but not limited to) those related to race and gender (Liberman et al., 2017; Rule & Sutherland, 2017). We also quickly assess their physical attractiveness (Ritchie et al., 2017). Attractiveness is a complex characteristic in that it relates not only to physical features such as facial symmetry and averageness (Thornhill & Gangestad, 1999), but also culture-specific associations of certain traits with high social status (Frederick et al., 2015). As such, perceptions of attractiveness are inherently linked to membership in salient social groups.

These quick inferences of strangers' traits are likely to color our subsequent impressions, potentially helping to explain well-documented patterns of ethnic and "lookist" bias. This makes it all the more surprising that little research to date has investigated the possible interplay of physical attractiveness and important axes of stratification such as race, ethnicity, and religion. What few examples of such research do exist highlight that studying the interplay of these two factors is not as simple as merely summing their effects. Rather, the effect of attractiveness seems to vary substantially across other social groups (Kunst et al., 2023; Monk et al., 2021; Ryabov, 2019).

In this dissertation, I will explore the intersection of physical attractiveness and group boundaries as it relates to three primary outcomes: wage setting, hiring, and trustworthiness perceptions. The first two studies focus on the labor market, a context in which physical attractiveness and ethnicity have been separately linked to persistent inequalities (Nault et al., 2020; Zschirnt & Ruedin, 2016). The final two studies focus on trustworthiness, a factor that has also been associated with attractiveness and ethnicity (Gereke et al., 2020; Wilson & Eckel, 2006), and which might mediate the link between appearance and labor market outcomes. Beyond the labor market, trustworthiness perceptions are important to a variety of outcomes as a key driver of social decision making and ultimately, social cohesion (Cook, 2001; van 't Wout & Sanfey, 2008). By studying the effects of appearance at the level of interpersonal beliefs, perceptions, and expectations, this dissertation aims to identify possible mechanisms that drive group-level inequalities. Applying the logic of the Coleman boat (Coleman, 1986), I study how macro-level stereotypes about appearance affect inter-individual perceptions. These perceptions may then affect economic

decisions like wage setting and hiring, which then yield economic inequalities. In doing so, I contribute to a deeper understanding of the micro-level drivers that are associated with macro-level appearance-based prejudice and discrimination.

Why study the nexus between appearance and discrimination? From a theoretical standpoint, appearance is interesting because it can signal so many qualities at once. Research in the tradition of status characteristics theory has long been interested in how various traits combine in producing patterns of inequality (Berger et al., 1992). In the last few decades, intersectionality theorists have also examined how the combination of different characteristics can often yield results that differ from the sum of their respective parts (Crenshaw, 1989; McCall, 2005). By studying the effect of visual cues, perhaps the most widely used signals of various traits in everyday life, this dissertation contributes to work on both status characteristics and intersectionality. While much sociological research treats social groups as monoliths and focuses only on between-group differences, this dissertation adds to a growing body of research that examines the potential causes of growing within-group inequalities (Leicht, 2008; Monk, 2022).

Perhaps more importantly, labor market discrimination has been stubbornly persistent over the last few decades (Quillian & Lee, 2023). This pattern of differential treatment contributes to intergroup economic inequalities and has detrimental effects on the psychological well-being of people in stigmatized groups (Darity, 2003; Jasinskaja-Lahti et al., 2007). This research aims to deepen our understanding of the drivers of stereotyping and discrimination, a necessary first step toward building a more equal society.

1.1 State of the art: appearance-based bias

I begin with a review of previous findings regarding the extent of ethnic, religious, and lookist biases as they relate to the two categories I study here: labor market outcomes and trustworthiness perceptions. I focus primarily (but not exclusively) on research in Europe, to provide the necessary context for my research conducted in Germany and Turkey. To conclude this section, I will also review the few studies that have previously examined the combinations of these forms of bias. While each form has separately attracted a great deal of scholarly attention, research on intersectional bias remains rare.

1.1.1 Ethno-religious bias

The studies presented here build on a wealth of research that has found evidence of discrimination against stigmatized ethnic groups. On the labor market, this discrimination takes many forms and spans the entire process from hiring to firing. To start, a great deal of field experimental research finds that people with migration background are less likely to receive invitations to interview when applying

for jobs (Lippens et al., 2023; Zschirnt & Ruedin, 2016). This overall result conceals considerable heterogeneity, as members of some groups face more difficulties than others. Applicants whose ancestry is traced to more culturally similar countries are preferred over those from more culturally distant settings (Ahmad, 2020; Friberg & Midtbøen, 2018; Vernby & Dancygier, 2019). Ethnicity may also interact with gender: among those with Middle Eastern or North African heritage, it seems that men face more disadvantages than women (Bursell, 2014; Dahl & Krog, 2018; Di Stasio & Larsen, 2020). Discrimination continues even after applicants pass the interview stage: even the few ethnic minority applicants who make it through the first stage are less likely to receive job offers (Quillian et al., 2020).

When people with migration background do find work in Europe, they are paid less than their ethnic majority peers, a disparity which is especially large for workers with non-European ancestry (Adsera & Chiswick, 2007; Ingwersen & Thomsen, 2021; Lehmer & Ludsteck, 2011). Ethnic minority men, and particularly those of non-European descent, are also more likely to be hired for unskilled work in precarious sectors, leaving them particularly vulnerable to economic shifts and making career advancement more difficult (Kogan, 2004, 2007). Finally, when an employer is forced to lay off workers, ethnic minority workers are more likely than their ethnic majority peers to be fired (Auer, 2022). In total, it is generally more difficult for people (and perhaps particularly men) with a migration background to get hired, advance in their careers, and stay employed over time.

Adherents of minority religious groups may face similarly discriminatory treatment in the labor market. Studies in the European context face the dilemma that most Muslims in Europe also have a migration background, possibly confounding the effects of religion and ethnicity. The few studies that have isolated the effect of religion have found that signaling adherence to Islam is associated with additional penalties in hiring, above and beyond any ethnic effect (Adida et al., 2010; Di Stasio et al., 2021; Valfort, 2020). These effects may be particularly strong for Muslim women who wear a veil, which makes religious practice more visible and salient to employers (Ahmed & Gorey, 2023; Fernández-Reino et al., 2023; Weichselbaumer, 2020). This particularly harsh treatment of veiled women may also stem from a general rejection of perceived religious fundamentalism, which is perceived to clash with secular European values (Helbling, 2014; Helbling et al., 2022).

Compared to the clear patterns found in labor market research, the connection between ethnicity and perceptions of trustworthiness is less clear. We would expect ethnic majority group members to find outgroup members less trustworthy, as suggested by some recent work (Schmid et al., 2022; Sofer et al., 2017). However, field experiments show conflicting results about the trustworthiness of immigrants and ethnic minority group members, suggesting that group-specific stereotypes in specific contexts may drive results (Bouckaert & Dhaene, 2004; Cox & Orman, 2015; Gereke & Ruedin, 2023; Kanitsar, 2023). These patterns may also intersect with other

traits, as men with a migration background might be seen as particularly untrustworthy (Gereke et al., 2020).

Religion, on the other hand, might have a positive effect on perceptions of trustworthiness. We tend to assume that religious people are more prosocial, perhaps owing to their presumed fear of divine retribution (Purzycki et al., 2016) or because of associations between religious practice and volunteering (Kelly et al., 2024). While religion is not an inherently visible characteristic, many people choose to wear religious badges, such as headscarves or jewelry featuring a cross or Star of David, to make their religious affiliation visible to observers (Sosis, 2005). Whether signaled through badges or by other means, religiosity seems to have a positive effect on perceived trustworthiness, even across lines of religious affiliation (Chuah et al., 2016; Power, 2017b; Tan & Vogel, 2008; Thunström et al., 2021). However, this effect seems to be modified by an observer's own level of religiosity, i.e., nonbelievers do not always trust the religious more (Thunström et al., 2021). This is particularly relevant to research in the German case, where a substantial portion of adults identify as nonreligious.

1.1.2 Beauty bias

The combination of physical attractiveness and ethnicity is particularly interesting because their effects tend to run in opposite directions: while members of stigmatized ethnic groups face substantial discrimination across many life domains, physically attractive people are broadly favored. As Dion et al. (1972) famously put it more than 50 years ago, we tend to believe that "what is beautiful is good." Accordingly, physically attractive people are thought to be more moral (Klebl et al., 2022), more prosocial (Hansson et al., 2024), and more intelligent than their less attractive peers (L. A. Jackson et al., 1995; Zebrowitz et al., 2002), among various other positive stereotypes (Eagly et al., 1991). Given this perceived connection between beauty and moral behavior, it is unsurprising that attractiveness has also been linked to higher perceived trustworthiness (Wilson & Eckel, 2006).

Perhaps owing to these myriad positive stereotypes, physical attractiveness is also rewarded on the labor market. Attractive people are more likely to receive positive responses when applying for jobs (Galarza & Yamada, 2014; Goulão et al., 2024; López Bóo et al., 2013; Maurer-Fazio & Lei, 2015; Ruffle & Shtudiner, 2015). Highly attractive people also earn higher salaries (Hamermesh & Biddle, 1994; Wong & Penner, 2016), and are more likely to be recommended for promotion (Ling et al., 2019; Morrow et al., 1990). A recent meta-analysis suggests that these advantages reflect employers' beliefs that attractive people possess more human and social capital (Nault et al., 2020).

However, there is also some evidence to suggest that the benefits of attractiveness are not universal. Gender in particular seems to play a strong moderating role with respect to the effect of attractiveness on labor market outcomes. Research has shown that men reliably benefit from attractiveness in a variety of contexts, while women

benefit only under certain conditions (Kukkonen et al., 2024). In some cases, highly attractive women may even face labor market penalties, a phenomenon dubbed the "beauty is beastly" effect (Heilman & Saruwatari, 1979; S. K. Johnson et al., 2010). This research provides some initial evidence that the value of attractiveness varies over important social categories.

1.1.3   Intersecting ethno-religious identity and beauty

What is less clear from existing research is how the effects of beauty might vary over categories defined by ethnic or religious identity. What little research exists regarding this interaction has focused primarily on labor market outcomes. One of these studies, a correspondence test from Peru, suggests that the effect of attractiveness on callback rates is greater for higher-status groups (i.e., White applicants compared to Indigenous applicants, Galarza & Yamada, 2014). However, research on race in the U.S. context finds the opposite result, showing evidence that Black women benefit the most from beauty, both in terms of wages and perceived suitability for employment (Kunst et al., 2023; Monk et al., 2021). These contradictory results suggest that this interaction might vary by geographic and social context, and no research to date has studied this question with respect to ethnic boundaries in Europe. This dissertation aims to fill this research gap, determining whether ethnic and religious identity might act as significant moderators of the beauty premium in Germany and Turkey.

1.2   Theory: at the intersection of beauty & identity

While each of the studies presented in this dissertation will expound their own theoretical arguments, I preface these works with some general theoretical considerations that guide my predictions about the intersection of ethno-religious identity and physical attractiveness. Generally, the work presented here tests the idea that ethno-religious identity might meaningfully moderate the beauty premium in terms of labor market outcomes and/or trustworthiness perceptions.

   This question may also rest upon the assumption that perceptions of attractiveness vary across groups defined by ethno-religious boundaries (Monk et al., 2021). Importantly, this does not suggest that members of some ethnic or religious groups are inherently more attractive than members of other groups. Rather, it suggests that prevailing cultural views of attractiveness within a given society are structured by local power relations. While there are certain facial features that are broadly considered attractive across cultures, our perceptions of attractiveness are also socially constructed (DeLamater & Hyde, 1998). For example, people whose appearance reflects aspirational standards of social privilege, i.e., those who most resemble members of high-status groups, are generally considered beautiful (Mears, 2014). Taking the perspective of dominant groups is especially important when considering labor market outcomes, as members of these groups are more likely to occupy high-status

positions that enable them to set wages and hire workers. Thus, I adopt a definition of attractiveness here that focuses on *perceptions* within a given society, rather than a purely "objective" measure of beauty (such as facial symmetry, etc.).

1.2.1 Unequal distribution of beauty

Social identity theory offers one theoretical explanation for the association between high-status groups and physical attractiveness (Tajfel & Turner, 1979). In this theoretical tradition, it is posited that people categorize others based on their membership in salient social groups (including but not limited to ethnicity, race, and gender). Those that share group memberships with the observer are viewed positively and those that do not are viewed negatively. Observers have an incentive to derogate outgroups because they gain esteem from belonging to a supposedly superior ingroup, especially when a particular outgroup is seen as a threat to the ingroup's position on the status hierarchy.

Attractiveness does not usually constitute an important part of one's social identity, and thus it is unlikely to function as a group boundary. It may however act as a medium for social comparison, or the valorization of one's ingroup relative to outgroups (Turner, 1975). People may seek to differentiate their ingroup from disliked outgroups on the basis of beauty: i.e., ingroup members are beautiful and outgroup members are ugly. An extreme form of this phenomenon can be found in racist and anti-Semitic caricatures, which tend to exaggerate physical differences between these outgroups and the ethnic or racial majority and depict members of these groups as unattractive (Ranta, 2017; Samples, 2019). Negative stereotypes about appearance may be one reason why so-called "sexual racism" persists in many societies today (Callander et al., 2015; Ranzini & Rosenbaum, 2020). Even less visible group boundaries also matter for perceptions of attractiveness: people tend to rate targets who share group memberships like political party affiliation or sexual orientation as more attractive (Nicholson et al., 2016; Schwartzman & Rule, 2024). While these results only suggest a pattern of in-group homophily and not an outright preference for high-status groups, the power imbalance between groups can lead to consensual views that members of ethnic minority groups are less attractive (Rudman & McLean, 2016). Thus, these results suggest that members of the high-status ethnic majority group should be seen as more attractive on average, while members of lower-status groups may suffer from appearance stigma (Goffman, 1963).

Literature on intersectional stereotyping offers another theoretical explanation for the unequal distribution of attractiveness. The theory of intersectionality suggests that stereotypes associated with race or ethnicity and gender are inextricably linked (Ghavami & Peplau, 2013; K. L. Johnson et al., 2012; Mize, 2024). Because sex typicality has been linked to attractiveness (Fiala et al., 2021), gendered racial stereotypes can affect the perceived attractiveness of ethnic minority people. Most research in this area has focused on race in the U.S. context, finding that prevailing cultural stereotypes of Asians as feminine leads to lower ratings of attractiveness for

Asian men (Wilkins et al., 2011), while stereotypes of African-Americans as masculine reduce the perceived attractiveness of Black women (M. E. Hill, 2002).

Translating these findings to Muslims in the European context is complicated by the fact that prevailing cultural stereotypes of Muslims differ starkly by gender. Muslim men are often stereotyped as threatening and dominant, while Muslim women are viewed as the submissive victims of their antisocial behavior (Fourgassie et al., 2024; Wiemers et al., 2024; Wigger, 2019). Thus, Muslim men (and by extension, men whose ancestry is traced to Muslim-majority countries like Turkey) might be seen as particularly masculine, and Muslim women might be seen as particularly feminine. These stereotypes may enhance perceived sex typicality and potentially also attractiveness.

In preliminary work done for this dissertation, I find some initial evidence for the paradigm suggested by work on intersectional stereotyping. The studies presented in Chapters 3-5 use face imagery from three academic face databases as stimuli (Ma et al., 2015; Saribay et al., 2018; Veit & Essien, 2022). To sort the initial selection of 108 photos by ethnic typicality and physical attractiveness, I had each photo rated by respondents drawn from an online access panel of 1,125 German residents on a variety of factors, including sex typicality and physical attractiveness. As presented in Figure 1.1, Turkish-heritage women were perceived as more feminine than White German women, and Turkish-heritage men were perceived as more masculine than White German men. For women, there is also a significant difference in ratings of physical attractiveness: Turkish-heritage women are slightly preferred. However, White German men and Turkish-heritage men were rated as roughly equally attractive. While this offers partial support for the intersectional stereotyping perspective (at least for women), it is also important to note that my selection of photos is not representative of the entire German and Turkish-origin population, and thus these results cannot necessarily be extrapolated to population-level differences.

Turning to religious rather than ethnic differences, similar theoretical explanations can be applied. In this dissertation, I signal religiosity using two cues common in the Muslim world: headscarves for women and beards for men. The effects
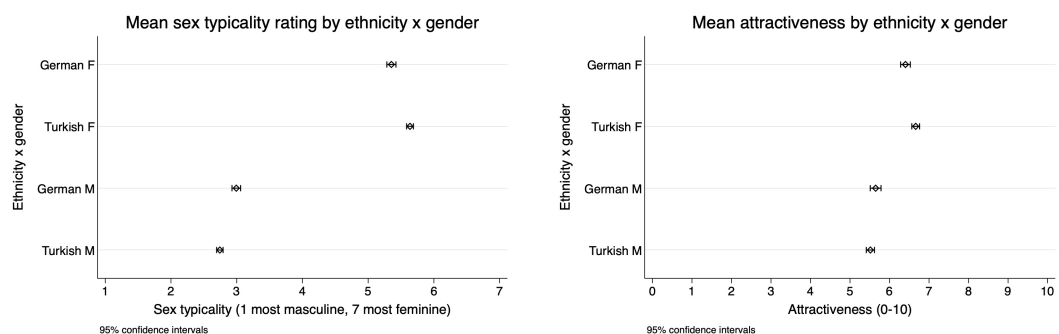
FIGURE 1.1: Ratings of sex typicality (1-7, left) and physical attractiveness (0-10, right) by phenotype and gender. $N = 1,125$ German residents.

of these religious badges are likely to vary across the two settings I study. In the German context, social identity theory may be particularly relevant, as headscarves are stigmatized as a visible signal of religious difference (Fernández-Reino et al., 2023; Weichselbaumer, 2020). In the Turkish context, where Islam is by far the most common religion, the headscarf may be less stigmatized as a religious signal. However, there are deep social cleavages regarding the role of religion in Turkish public life, and wearing a veil may still be stigmatized by more secular Turks as a sign of the wearer's (presumably conservative) political beliefs (Saktanber & Çorbacioğlu, 2008). It is also important to note that headscarves necessarily cover the hair and parts of the face and neck, giving observers less information on which to rate the wearer's attractiveness. This simple explanation may explain findings from Iran and the United Arab Emirates that suggest the headscarf is associated with lower ratings of attractiveness, even among Muslims in these more conservative settings (Jordan et al., 2020; Pazhoohi & Hosseinchari, 2014; Sheen et al., 2018). For men, research on beards suggests that they do not have a clear effect on perceptions of attractiveness (B. J. Dixson & Vasey, 2012), although perceptions seem to vary across contexts (Garza et al., 2023).

1.2.2 Unequal rewards of beauty

The unequal distribution of attractiveness across social groups may also influence the relative value of attractiveness for members of these groups. Sociologists have conceived of physical attractiveness as a form of capital, often called bodily capital or erotic capital (Bourdieu, 1984; Hakim, 2010), which can be exchanged for other forms of capital, such as earnings. Monk et al. (2021), building on the theory of Bourdieu (1986), suggest that the value of this capital is set by its scarcity within a given group. Therefore, members of groups with a generally stigmatized appearance (such as Black Americans) who are nonetheless seen as attractive should receive an even larger benefit from their bodily capital for standing out from the other members of their group.

Other theoretical frameworks would also predict a pattern of larger returns for members of stigmatized groups. Physical attractiveness has also been conceptualized as a diffuse status characteristic, or a socially salient trait with defined states that differ in terms of their perceived status (Webster & Driskell, 1983). According to status characteristics theory (SCT), people who have high-status traits such as high physical attractiveness are deemed to be more competent than those with lower-status traits (Berger et al., 1972). This effect would explain the association of beauty with a variety of positive characteristics, even those that seem *prima facie* unrelated to appearance (Eagly et al., 1991; L. A. Jackson et al., 1995).

Status characteristics theory posits that various status characteristics (such as beauty, gender, race, age, etc.) are combined in a generally additive fashion to yield a sum total of a person's status (Webster & Driskell, 1978). In this way, high-status characteristics can counteract the effect of low-status characteristics, and vice versa.

However, each additional characteristic with the same level of status has a decreasing marginal effect (Berger et al., 1992; Pedulla, 2018). This suggests that attractiveness, a high-status characteristic, would have a smaller effect for people who already have many other high-status characteristics (such as ethnic majority German men) and a larger effect for those with few high-status characteristics (such as women with a migration background).

It has historically been assumed that stereotyping works in a similarly additive fashion, with positively stereotyped traits canceling out the effects of negatively stereotyped traits (Pedulla, 2018). However, recent work has complicated this picture by taking an intersectional approach to stereotype research, assuming that stereotype content varies at the intersection of race or ethnicity and gender, as well as other important social categories like social class (Benard et al., 2023; Heiserman, 2023; Mize, 2024). This work suggests that stereotype content does vary across intersecting social categories, but more work will be needed to identify patterns in these combinations. Thus, I expect that stereotypes will differ across ethnic and gender groups, but prior work does not yet facilitate an easy prediction about the patterns of difference that I may find.

Finally, returning to social identity theory would lead me to an opposing prediction about the relative value of attractiveness across ethnic groups. Building on the general idea that groups might seek positive distinctiveness by portraying their members as more attractive than members of outgroups, the "black sheep effect" suggests that attractiveness would produce larger rewards for attractive ingroup members and harsher penalties for unattractive ingroup members (Marques et al., 1988). These penalties at least symbolically "purge" undesirable group members in order to maintain the overall positive assessment of one's ingroup (Marques & Paez, 1994). Similarly, outgroup members would be rewarded less for their attractiveness in order to maintain clear separation between groups on the basis of beauty. Applied to the present research question, this framework would predict larger beauty premia and larger plainness penalties for ethnic majority Germans, at least from the perspective of mostly ethnic majority decision makers. If this pattern holds, physical attractiveness should matter much more for the ethnic majority than for members of ethnic minority groups.

To sum up, while both Bourdieusian theories of capital and status characteristics theory would predict a larger beauty premium for members of low-status groups, social identity theory would predict the opposite. Additionally, intersectionality theory and the stereotype content model would predict that stereotypes vary over groups defined by salient social categories, but these perspectives offer little predictive guidance about the relative effect of attractiveness in each group. The following studies will offer new insights into patterns of discrimination in these categories, helping to resolve the contradictions of these commonly applied social theories.

1.3 Contribution: appearance and identity in two contexts

1.3.1 Research questions

In this dissertation, I present findings from original research that was designed to answer several research questions. However, all of these questions revolve around one broader question: how do the effects of appearance vary over groups defined by ethnicity, religiosity, and gender? Building on the work of status characteristics and intersectionality theories, and reflecting on the persistent ethnic and religious bias found in European societies, I predict that physical attractiveness will modify religious and ethnic stereotypes. Ultimately, this may reduce the extent of ethno-religious discrimination against highly attractive people belonging to stigmatized groups. I also study the effects of religious badges as a second visual cue, which I expect to have an impact on perceptions of trustworthiness. Combining these questions, this central research question can be rephrased as follows:

Can appearance-based signals narrow ethnic and religious gaps in labor market outcomes and trustworthiness perceptions?

In the first half of this dissertation, I will focus on ethnic and religious gaps in labor market outcomes. More specifically, chapter 2 explores the effects of ethnicity and attractiveness on wages in the German labor market. Here, I ask whether physical attractiveness can help close persistent immigrant wage gaps, and whether this effect is also moderated by the degree of cultural similarity between the host and heritage countries of workers with a migration background.

Next, I turn my attention to disparities in hiring. Chapter 3 tests the intersection of several highly salient traits on callback rates, including gender, ethnicity, religion, and physical attractiveness. Here too, I ask whether physical attractiveness has a larger effect for members of stigmatized groups, helping to close ethnic gaps in callback rates. I also explore how other factors, such as information about work performance and occupational characteristics, might interact with physical attractiveness with respect to rates of positive response.

The second half of the dissertation instead focuses on how appearance-related characteristics affect perceptions of trustworthiness, exploring the role of factors like attractiveness, ethnicity, and religiosity. In Chapter 4, I test the combined effects of physical attractiveness and ethnicity on the perceived trustworthiness of ethnic majority German and Turkish-background men. Finally, in Chapter 5, I test the effects of religious badges in the Muslim-majority setting of Turkey, exploring how religiosity (and in some analyses, also attractiveness) influences perceptions of trustworthiness.

1.3.2   Overview

In the following chapters of this dissertation, I present results from four studies that investigate the effects of appearance and ethno-religious identity on three outcome variables in two geographic contexts, as shown in Table 1.1. The sole non-experimental chapter (Chapter 2) analyzes a large, longitudinal panel dataset to explore the "real-world" consequences of appearance on earnings. Here, we employ a multiverse analysis, estimating hundreds of model specifications to test the robustness of our findings.

Study	Context	Independent variables				Dependent variables		
		Physical attractiveness	Ethnicity	Gender	Religiosity	Wages	Hiring	Trustworthiness perceptions
Chapter 2	Germany	x	x	x		x		
Chapter 3	Germany	x	x	x	x		x	
Chapter 4	Germany	x	x					x
Chapter 5	Turkey	x		x	x			x

TABLE 1.1: Independent variables, dependent variables, and context, by study.

The other three studies in this dissertation use experimental methods to estimate the causal effects of appearance-related characteristics on these outcomes of interest. These methods allow me to randomly assign a variety of characteristics to vignette persons, such as ethnicity, religion, gender, and physical attractiveness, and evaluate how different combinations of these traits are perceived. Because this dissertation is focused on appearance-related bias, I employ primarily visual cues, manipulating the vignette person's identity through the use of pretested photo stimuli.

Chapter 3 employs a randomized field experimental approach, long considered the gold standard in measuring discrimination (Gaddis, 2018). Chapters 4 and 5 use vignette experiments to test how respondents perceive profiles that vary in terms of their name and appearance. These vignette studies also allow me to evaluate the extent to which respondents' own characteristics influence their perceptions of the vignette persons, an important test of possible heterogeneity in my overall findings.

It is important to note that the operationalization of physical attractiveness used in these studies relates only to facial attractiveness, rather than an assessment of a person's whole body. Both the body and face have strong effects on perceptions of attractiveness, and may each signal different traits (Alicke et al., 1986). While I often use the terms "physical attractiveness" and "facial attractiveness" interchangeably (in keeping with the literature on the topic), the present work does not include variation in bodily features such as weight, which may also have effects on interpersonal

perceptions (Goulão et al., 2024; Rooth, 2009). This is a limitation that should be addressed in future studies.

Summary: Chapter 2

In the first paper of my dissertation, my co-authors and I test whether the economic returns of physical attractiveness vary according to immigrant status and gender. A wealth of research has already established that female workers and those with migration background face wage penalties in the German labor market, but no work to date has investigated whether these inequities are also affected by physical attractiveness. We expect that ethnic majority Germans and people with migration background will differ not only in terms of the amount of "bodily capital" they possess, but also their ability to transform that capital into earnings.

It is however unclear *a priori* how ethnicity will affect the size of the beauty premium in the German context. We derive two sets of competing hypotheses from two strands of literature. Research on resource substitution theory would support the idea that the lowest-status groups (i.e., people with migration background, and especially those from more culturally distant settings) would receive the largest benefit from attractiveness. On the other hand, the human capital perspective would suggest that people with migration background may not be able to transfer their capital as easily outside of their heritage country, leading to a larger premium for ethnic majority natives.

For this analysis, we use longitudinal data from the German Family Panel pairfam (Brüderl et al., 2021). The pairfam dataset is uniquely well-suited to our analysis, offering interviewer ratings of respondents' attractiveness as well as information about their ethnicity and their earnings over time. We use this information to estimate traditional panel regression models of the effect of attractiveness on earnings, as well as a multiverse analysis which tests the robustness of our findings to alternate specifications. Across 864 model specifications, we find evidence that migration background does seem to moderate the effect of attractiveness on earnings. Specifically, we find evidence for a substantial beauty premium in earnings among native German men and women, but results for respondents belonging to ethnic minority groups are inconsistent. Some models show a beauty premium for women with a Turkish migration background, suggesting a similar pattern to that found among Black American women by Monk et al. (2021), but these effects are only significant in about half of our models. Overall, these findings provide some initial evidence that the size of the beauty premium is moderated by ethnicity, a proof of concept for the project that leads us to investigate whether the same patterns exist for other outcomes.

Summary: Chapter 3

Specifically, the next project examines whether the patterns we find for earnings also apply to hiring behavior in the German labor market. Previous work has established that people with migration background receive fewer callbacks when applying for jobs in Europe (Zschirnt & Ruedin, 2016). A smaller body of work has also found that employers are more likely to offer interviews to more attractive applicants (Nault et al., 2020), although this has not yet been tested in Germany. Here, we aim to explore the interplay between these two factors and assess whether attractiveness may help close ethnic gaps in hiring.

In this study, we conduct a randomized correspondence test in the German labor market to test the causal effects of ethnicity, gender, and physical attractiveness on callbacks in eight occupations. We send more than 3,800 fictitious résumés to real job openings throughout the country, exploiting the German custom of including a photo with a job application to signal attractiveness using a pre-rated headshot. By comparing response rates across categories defined by ethnicity, gender, religiosity, and physical attractiveness, we can determine whether there are ethnic penalties, "plainness penalties," and whether these two factors interact with respect to callback rates.

Our results show evidence of an ethnic hierarchy: applicants with Turkish migration background face significant discrimination, as do men with Greek migration background, while Danish-heritage applicants face no discrimination relative to the ethnic majority. While there is a modest beauty premium in overall callback rates, this premium does not narrow the German-Turkish ethnic gap. If anything, beauty offers significant benefits to ethnic majority German women, widening the gap between the ethnic majority and ethnic minority applicants at higher levels of attractiveness. This study contributes to the literature on labor market discrimination and status characteristics, illustrating how seemingly advantageous traits can reinforce ethnic stratification rather than mitigate it.

Summary: Chapter 4

Echoing the design of my labor market correspondence test, I first examine the interplay of ethnicity and attractiveness in forming perceptions of trustworthiness. While previous research has explored the impact of these two factors separately, finding ambiguous results about the trustworthiness of immigrants (Gereke & Ruedin, 2023; Kanitsar, 2023) but a positive association between trustworthiness and attractiveness (Wilson & Eckel, 2006), this study is the first to explore the intersection of these two highly salient characteristics with respect to trustworthiness.

To assess whether the size of the beauty premium in trustworthiness perceptions varies across ethnic groups, I use a vignette experiment embedded within wave 70 the German Internet Panel (GIP), a longitudinal study with a large sample that is representative of the German population (Blom et al., 2015). As a measure of the

perceived trustworthiness of a vignette person, I use the "lost wallet question," a tool previously used to measure relational trust (Soroka et al., 2007). Each of the 1,794 ethnic majority German respondents randomly selected for this experiment were presented with one lost wallet vignette containing an ethnically typical name and a headshot-style photo of the person described, and asked to rate the trustworthiness of the vignette person.

The results of this survey experiment largely defied my theoretical expectations. While I do find that respondents perceive more attractive vignette persons to be more trustworthy, as expected, respondents reported that Turkish-origin vignettes were seen as *more* trustworthy than ethnic majority vignettes. Finally, the size of the beauty premium does not vary between the two ethnic groups, although the effect of attractiveness is only significant (and positive) for Turkish-heritage vignette persons. To help explain this surprising result, I performed several exploratory analyses in which I divide the sample into subgroups. Some of the most dramatic differences can be seen between respondents with inclusionary immigration attitudes and exclusionary immigration attitudes. While more exclusionary respondents do not differentiate substantially between Turkish-origin and ethnic majority vignette persons, more inclusionary respondents report finding the Mehmet vignettes more trustworthy, thus driving the overall results. These results suggest that while facial attractiveness does not significantly moderate trustworthiness perceptions, respondents' immigration attitudes are highly relevant.

Summary: Chapter 5

Finally, we turn our attention to the effects of religiosity, a trait that should theoretically be more relevant to perceptions of trustworthiness than ethnicity. There are several reasons to expect that highly religious people may be more trustworthy than their less religious peers. First, religious people generally believe that their actions are being watched by a supernatural entity and as a result, they may fear divine retribution for betraying others' trust (Norenzayan, 2014; Purzycki et al., 2016). Second, religious people may be afforded more opportunities to engage in prosocial behavior due to the involvement of religious communities in charitable projects (Kelly et al., 2024). Finally, individuals who are inherently more prosocial might seek out religious communities precisely because they wish to engage in these activities (Aksoy & Wiertz, 2024).

To study how visible religious signals are perceived, we conducted a face-to-face survey of 2,170 Turkish adults. Surveyors presented each participant with six vignettes and marked their responses by hand in a printed booklet. Each vignette consists of a photograph of the vignette person and a short description. The photo conveys the person's gender, their level of religiosity, and their level of physical attractiveness (low or high), and the text indicates one of six common occupations which act as an indication of social class. We convey three levels of religiosity by editing our set of eight photos to add visual markers of religiosity that are common

in the Turkish context: head coverings for women and facial hair for men. As in the previous study, we measure trustworthiness perceptions by asking a "lost wallet question" about the likelihood of the vignette person returning the respondent's wallet if it were lost in a public place. We hypothesize that vignettes wearing religious signals will be seen as more trustworthy, an effect that should be stronger among more religious respondents. We also expect that more attractive vignette persons will be seen as more trustworthy, in line with previous studies.

Contrary to these expectations, we find that religious signals are not significantly associated with higher perceived trustworthiness. Women wearing a headscarf are seen as roughly equally trustworthy as secular women, while all other religious and devout vignette persons are thought to be significantly less trustworthy than secular vignette persons. Respondent religiosity plays a significant role in perceptions, as more religious respondents rate religious vignette persons as more trustworthy than their less religious peers do. More religious respondents even find women wearing a headscarf to be significantly more trustworthy than women in the secular condition. Finally, we do find a trust premium for more attractive vignettes in most categories, although differences between attractiveness categories are no longer significant for women in the religious and devout conditions. This suggests that covering a woman's hair lessens the influence of facial attractiveness on trustworthiness perceptions. These findings suggest that contrary to a wealth of evidence, religiosity is not universally linked with increased trustworthiness. Rather, the link between religiosity and perceptions of trustworthiness seems to depend on local context.

Chapter 2

Pretty Unequal? Immigrant-Native Differences in Returns to Physical Attractiveness in Germany

Abstract

How do the economic returns to physical attractiveness vary between immigrants and natives? Despite a growing literature on the beauty premium, previous scholarship has not examined whether the size of this premium might vary according to immigrant status, an important axis of social stratification in European societies. We study this question in the German labor market using longitudinal data from a large, randomly drawn nationwide sample that includes interviewer ratings of respondents' physical attractiveness. Results from a multiverse analysis provide strong evidence of a beauty premium for native German men and women. However, we find mixed results for ethnic minority groups. These results highlight how perceived physical attractiveness is linked to labor market stratification and suggest that physical attractiveness may be one mechanism driving the persistent immigrant-native wage gap.

2.1   Introduction

It has been fifty years since the "what is beautiful is good" hypothesis was first proposed, implying that physical attractiveness has powerful positive effects on how a person is perceived (Dion et al., 1972). Since then, it has become increasingly clear that these perceptions are linked to preferential treatment toward attractive people, and that these advantages can accumulate over the life course into significant inequalities across a variety of domains (Gordon et al., 2013; Hamermesh, 2011; Jæger, 2011). Among this wide array of benefits, one of the most commonly studied has been the "beauty premium," or the increase in earnings associated with physical

---

attractiveness (Deryugina & Shurchkov, 2015; Doorley & Sierminska, 2015; Hamermesh & Biddle, 1994; Scholz & Sicinski, 2015; Shapir & Shtudiner, 2022). However, this broad finding may conceal heterogeneous effects. For example, the effect of attractiveness on earnings may vary by gender, with some studies finding larger effects for men (Doorley & Sierminska, 2015) and others finding larger effects for women (French, 2002). Additionally, a recent study on the U.S. labor market finds that race and gender both influence the size of the beauty premium, and specifically that historically disadvantaged groups receive the largest benefit (Monk et al., 2021). These results suggest that the value of physical attractiveness is "inextricably linked to the 'value' of its bearer" (Monk et al., 2021, p. 204) and that the beauty premium is likely structured by the primary axes of social division in a given society. While race is a highly salient boundary in the U.S., in the European context, ethnicity and immigrant status are more frequently discussed (Alba, 2005; Gereke et al., 2022).

In this paper, we test whether the economic returns of physical attractiveness might vary according to immigrant status. Bringing together two bodies of literature, we derive competing theoretical expectations. The first, the "resource substitution" perspective, suggests that any given resource is more valuable to people who have fewer alternative resources at their disposal (Ross & Mirowsky, 2006). From this perspective, we would expect that attractiveness would have the largest effect on earnings for members of marginalized groups, like immigrants and people belonging to an ethnic minority. However, research on immigrants' human capital (the "human capital" perspective) suggests that they would be less able to transfer their capital into earnings if it was obtained outside of their country of residence (Friedberg, 2000). While bodily capital is not "obtained" in a specific location like education or work experience, its value may be culturally specific due to variation in beauty standards across societies. Thus, immigrants, and especially those whose families originate from more culturally distant settings, may receive a smaller beauty premium than majority-born natives.

To study the interplay of physical attractiveness and ethnicity (with respect to immigrant status) in relation to earnings, we use longitudinal data from the German Family Panel pairfam (Brüderl et al., 2021). Importantly, this dataset includes interviewer ratings of the physical attractiveness of each respondent in the first wave, as well as data about each respondent's ethnic background, and their earnings over time. In addition to results from traditional panel regression models, we also present results from a multiverse analysis, which estimates plausible alternative specifications to test the robustness of our findings (Steegen et al., 2016).

Our research makes several advances to a growing body of research on the beauty premium and labor market stratification. Empirically, we provide the first evidence that ethnicity possibly moderates the effect of physical attractiveness on earnings. These results help to better understand how differences in perceived attractiveness may impact the much-discussed immigrant wage gap. If highly attractive natives are privileged in the labor market while highly attractive immigrants are not, this

suggests that physical attractiveness may be one mechanism driving the persistent immigrant wage gap in Germany. Additionally, we contribute methodologically to research on the beauty premium by applying multiverse methods. This method accounts for our degrees of freedom as researchers, providing robust evidence for the economic rewards of physical attractiveness. Running more than 2,500 models, we find strong evidence of a beauty premium for native German men and women. However, we find mixed evidence for immigrants and their descendants.

2.2 Immigrant-native wage inequalities and physical attractiveness in Germany

A considerable body of research finds substantial earning differences between immigrants and their descendants on the one hand and native Germans on the other. Depending on their country of origin, immigrants are affected to different degrees by wage penalties, with wage gaps between natives and immigrants increasing for those born in non-EU countries (Adsera & Chiswick, 2007). The wages of Southern European immigrants seem to differ less from those of their German counterparts than those of Polish, Turkish, or Ethnic German immigrants (Ingwersen & Thomsen, 2021; Lehmer & Ludsteck, 2011)[1]. It is also well established that not only immigrants but also women suffer from wage inequalities in Germany: women still earn substantially less than equally qualified men (Bonaccolto-Töpfer et al., 2023). These findings raise the possibility that foreign female workers may suffer from "double disadvantage," or discrimination both on the basis of their ethnicity and their gender (Piazzalunga, 2015).

In this article, we are interested in another essential factor for immigrants' labor market outcomes: their perceived physical attractiveness. Empirical studies have documented that physical attractiveness has positive consequences in the labor market, increasing wages as well as the likelihood of being hired or promoted (Hamermesh & Biddle, 1994; Judge et al., 2009; Ling et al., 2019; Wong & Penner, 2016). However, results regarding the question of whether men and women benefit equally from attractiveness are ambiguous. While some studies find that attractiveness has stronger effects for men than for women (Doorley & Sierminska, 2015; Ruffle & Shtudiner, 2015), other analyses indicate that women benefit more from physical attractiveness (Jæger, 2011; Maestripieri et al., 2017). Kukkonen et al. (2024) provide an explanation for these inconsistent results by arguing that both women and men benefit from physical attractiveness, but for women, it seems to be more

---

[1]The term "Ethnic German immigrants" (in German, "*Aussiedler*") refers to people of German ancestry who lived in Eastern Europe and returned to Germany in the second half of the 20th century, as well as their descendants. These are largely people who remained in formerly German territories, or people who fled persecution during World War II.

context-dependent. S. K. Johnson et al. (2010) report, for instance, that the occupational context matters and that very attractive female applicants face disadvantages in masculine sex-typed occupations in which perceived physical appearance is unimportant. Attractiveness may also affect men and women differently: research on the earnings penalty associated with obesity (a factor often connected to attractiveness) finds that overweight women are affected by taste-based discrimination while overweight men are disadvantaged by differences in human capital (Bozoyan & Wolbring, 2018).

Immigrants are affected by the beauty premium insofar as there is evidence that they are perceived as being less physically attractive by natives. For example, while European immigrants may be evaluated similarly to natives (Kočnar et al., 2019), non-European immigrants are generally less likely to have the characteristics that are most valued by Eurocentric beauty standards. Although there is not much evidence yet on the perceived attractiveness of immigrants in the European context, Ranzini and Rosenbaum (2020) show that Dutch adults rate Caucasian Tinder users as more attractive compared to Turkish, Moroccan, or Antillean Tinder users. U.S. research also suggests that skin color is likely to have a particularly significant effect on attractiveness perceptions (M. E. Hill, 2002)[2]. These findings suggest that people with a "foreign" appearance may be judged more negatively by others, an effect that may be even stronger for non-White people.

Not only do we expect that immigrants and natives will differ in terms of the amount of bodily capital that they possess, but we also expect that the returns on that capital may vary according to immigrant status (i.e., between immigrants and majority group natives). That is, a person's ability to turn their bodily capital into economic capital may depend on their identity, or more specifically their membership in stigmatized social groups (Monk et al., 2021). Thus, in this article, we are not only interested in the association of physical attractiveness and wages but also in how it varies at the intersection of gender and ethnicity. In doing so, we add to research that helps explain widening gaps in within-group inequalities (Leicht, 2008).

What remains unclear is how these returns will vary, especially in the German context. While Monk et al. (2021) find that Black Americans, and especially Black women, receive the greatest returns to attractiveness in the U.S., these results may not be generalizable to Germany, or to groups defined by immigrant status rather than race. Group boundaries in Germany differ substantially from those in the U.S. in that the primary axes of division are nationality, descent, and religion, while race is seldom discussed (Zolberg & Woon, 1999). These boundaries are generally less visible and thus potentially more malleable, leading to potential differences in how

---

[2]While results from Ranzini and Rosenbaum (2020) suggest that race may be relevant to perceptions of attractiveness in Europe, data about race or skin color is rarely collected in Germany. Thus, we are unable to assess the effect of race on the beauty premium despite the fact that it may represent a source of heterogeneity within the populations studied here.

appearance is evaluated (Gereke et al., 2022). Additionally, the German context differs in important aspects from the U.S., where most of the studies on beauty premia have been conducted. First, the German labor market is more regulated than its counterparts in liberal welfare states, like the U.S. Germany can be characterized as a coordinated market economy (P. A. Hall & Soskice, 2001), with comparatively strong unions and centralized collective bargaining, giving employers less discretion in wage setting. Second, the German education system is standardized, with a high degree of vocational specificity (Allmendinger, 1989). As such, educational credentials are relatively reliable signals for employee skills and productivity and reliance on additional signals, like attractiveness, might be less pronounced.

2.3 Competing theoretical perspectives

Previous literature on discrimination focuses on two primary mechanisms explaining preferential treatment in market settings: taste-based and statistical discrimination. Taste-based discrimination refers to the preference of an employer for a certain characteristic (such as beauty or a specific ethnic or racial identity) regardless of whether this characteristic is thought to make an applicant more productive (Becker, 1971). Statistical discrimination, on the other hand, refers to an employer using visible characteristics of a person to infer their productivity or fit for a position due to a lack of reliable information (Arrow, 1973; Phelps, 1972). An extensive body of work attempts to ascribe one of these motives to various forms of discrimination, or to examine the possible interplay between the two (Neilson & Ying, 2016). With regard to "lookist," or attractiveness-based discrimination on the labor market, a recent review finds more evidence for statistical discrimination, suggesting that employers expect that perceived beauty is associated with productivity, perhaps through traits like confidence or persuasiveness (Nault et al., 2020). However, there is more evidence for taste-based discrimination when it comes to ethnic discrimination, suggesting that preferential treatment for majority group members is not driven by a lack of information (Lippens et al., 2022). While these distinctions are important in understanding the mechanisms behind observed patterns of discrimination, we are unable to disentangle these mechanisms with the observational data we use here.

Therefore, we turn to other theoretical perspectives to explain the potential for ethnic variation in the beauty premium, leading us to contradictory hypotheses. One such perspective, the *resource substitution theory*, would support the findings of Monk et al. (2021), positing that members of stigmatized groups (such as certain ethnic minority groups) should reap the greatest rewards for physical attractiveness. This theory suggests that the effect of any given resource is larger for people who have fewer alternative resources at their disposal (Ross & Mirowsky, 2006). In other words, someone who has only one form of capital will be wholly reliant on it for their well-being, while someone with multiple forms of capital could use

any of them to increase their earnings, thus decreasing the marginal benefit of additional resources. Prior research has already applied this theory to bodily capital: a U.S. study finds that attractiveness has a larger effect on educational attainment for people with lower socio-economic status even though low-SES individuals are less likely to be seen as attractive (Bauldry et al., 2016). Recent work on German football players also suggests a substitution mechanism: lower-performing players receive a larger beauty premium (Chan et al., 2022). From this perspective, we would expect attractiveness to have the largest effect for people who belong to disadvantaged social groups, such as migrants.

A competing theoretical framework, the *human capital* perspective, would instead suggest that immigrants should be less able to translate their capital into earnings outside of their home country, thus reducing the reward they receive from physical attractiveness. Work in this field has generally focused on the transferability of human capital like work experience and education, finding that immigrants' foreign human capital is less valuable than human capital acquired locally (Basilio et al., 2017; Chiswick & Miller, 2009; Friedberg, 2000). While bodily capital differs from human capital in that it is not "acquired" in a specific location, we would expect its value to be tethered to local beauty standards, thus rendering it culturally specific. For example, in the same way that an employer might be unsure about the value of a foreign degree, employers may not be sure that an attractive but foreign-looking person will be able to use their attractiveness in the same way that an attractive native-looking person could. Employers may assume that their customers or coworkers would prefer to interact with natives (Baert & De Pauw, 2014), and thus disregard attractiveness for applicants that do not appear native[3]. This tendency might make attractiveness irrelevant for those who do not appear to be members of the native majority group, regardless of their actual immigrant status. Considering attractiveness from this perspective, we would expect that immigrants, and especially those who are phenotypically distinct from the majority population of native Germans, would receive a smaller beauty premium.

2.4   Hypotheses

From these two theoretical perspectives we derive several hypotheses, including two general hypotheses about the relative effect of attractiveness on earnings for immigrants and natives, as well as two additional hypotheses relating to subgroup analyses based on the dimension of ethnicity. While the mechanisms linking attractiveness and earnings seem to differ by gender, we refrain from making explicit

---

[3]This tendency may be especially strong in fields like sales or hospitality where employees have extensive contact with customers. Previous research finds that the beauty premium is especially large in these contexts (Deryugina & Shurchkov, 2015). While this represents a worthy topic for future research, we cannot investigate it here as our sample of ethnic minority respondents is not sufficiently large to examine interactions with the degree of customer contact or industry.

hypotheses about this, as it is not yet clear how gender specifically affects the link between attractiveness and earnings.

The resource substitution perspective suggests that people with fewer resources will benefit more from attractiveness. Given that immigrants have less human capital on average than natives, we would predict that they should benefit more from attractiveness in terms of earnings, as stated in Hypothesis 1A.

H1A: Immigrants will receive a greater reward from attractiveness than natives.

To extend this argument to differences between ethnic subgroups, we consider which groups are more or less similar to German natives. We expect that members of more culturally similar groups will face fewer obstacles in the labor market (Ingwersen & Thomsen, 2021). We would also expect that immigrants who come from more culturally distant environments would be less phenotypically similar to natives and thus more easily categorized as "non-Germans," which may negatively impact their labor market outcomes. Indeed, recent studies indicate that immigrants from culturally dissimilar countries face substantial discrimination while applying for jobs, while those from similar countries are treated like natives (Koopmans et al., 2019; Veit & Thijsen, 2021). This leads us to Hypothesis 2A, which suggests that precisely the groups that are likely to face the greatest discrimination should see the greatest reward from attractiveness.

H2A: Immigrants with the greatest cultural difference from natives will receive a greater reward for attractiveness than those with less cultural difference.

From the human capital perspective we derive contrary hypotheses. As a general hypothesis, we would expect that immigrants and their descendants would be less able to transfer their bodily capital into earnings than natives due to culturally specific beauty standards, leading us to Hypothesis 1B.

H1B: Immigrants will receive a smaller reward from attractiveness than natives.

Turning our attention to interethnic differences, we suppose that immigrants with greater cultural similarity to natives will be able to transfer their capital more easily (Ingwersen & Thomsen, 2021). We extend this line of thinking to physical appearance, assuming that a more similar appearance will also be easier to transfer across cultural boundaries as it will be more readily seen as "native."

H2B: Immigrants with the greatest cultural similarity to natives will receive a greater reward for attractiveness than those with more cultural difference.

2.5 Data & methods

Instead of estimating a single set of models with additional robustness checks, we conducted a so-called multiverse analysis (Simonsohn et al., 2020; Steegen et al.,

2016). Any single analysis is the result of a dizzying array of choices including variable coding, the inclusion of certain covariates, exclusion criteria, and model selection, some of which may be made without careful consideration of possible alternatives. Even when these choices are justified, different researchers may make different decisions, yielding substantively different results (Schweinsberg et al., 2021). To assess whether our results are a mere artifact of our design choices or if they constitute a true effect, we will estimate a variety of models for each gender and ethnicity pair, and present the effect sizes from each model in specification curves (Simonsohn et al., 2020). This aims to combat selective reporting by showing results from reasonable alternative model specifications in parallel (Steegen et al., 2016). The set of models we would consider reasonable for our research question is outlined in Table 1, which outlines the "garden of forking paths" we faced as we planned our analysis (Gelman & Hill, 2007). While we have noted the design decisions we feel are most suitable for our analysis (in italics), we estimate the model with each unique combination of the specifications listed, for a total of 2,592 models. In the interest of transparency, we also include decisions for which there is only one feasible option.

2.5.1 Data

To explore the possibility of ethnic variation in the beauty premium in Germany, we use data from waves 1-12 of the German Family Panel pairfam, release 12.0 (Brüderl et al., 2021). In 2008, the pairfam study began conducting annual in-person interviews with more than 12,000 randomly selected Germans from three birth cohorts (1971-73, 1981-83, and 1991-93) as well as their partners, parents, and children, with a focus on understanding partnership formation and family living arrangements in Germany (for more information, see Huinink et al. (2011))[4]. The pairfam sample includes an oversample of respondents of Turkish origin and Ethnic German immigrants from Eastern Europe and former Soviet states, making it a useful source of data to study ethnic inequalities in labor market outcomes.

2.5.2 Outcome variable: wages

As the outcome variable, we study respondents' hourly wage. Respondents are asked to indicate their individual earnings from the previous month. They are asked to exclude any extra earnings like vacation pay or back pay, but to include overtime pay. Thus, the amount reported may not accurately represent the amount earned in a typical month if the respondent worked much more than usual. Respondents also report the number of hours they work in an average week, which we use to calculate an average hourly wage (monthly earnings divided by the hours worked per week multiplied by 4.3). Since earnings are consistently positive and a multiplicative model on the original scale seems reasonable as we estimate a Mincer-type model

---

[4]While the study also includes a separate subsample of Germans from the eastern states (the "DemoDiff" sample), we had to exclude these due to missing information on our main independent variable in wave 1.

(Mincer, 1974), we use the natural logarithm of the hourly wage in some models. We use gross and net hourly wages in our multiverse analysis, but prefer the former as gross wages provide a better indicator of labor market rewards since they are a measure of pre-tax and pre-transfer earnings. However, pairfam only collects data on gross hourly earnings every two years, therefore we also use net hourly earnings in order to increase the sample size. Upon inspection of the wage data, we found that some respondents reported implausibly low or high earnings given their occupation and work hours. This may be due to errors like respondents reporting no work hours if they were ill or on vacation, or reporting annual rather than monthly weekly earnings. To address this, we estimate models with three versions of treatments of the outcome variable: unedited, including these improbable values; Winsorized, setting values below the 5th percentile or above the 99th percentile equal to the 5th and 99th percentiles, respectively; or trimmed, in which we exclude respondents who report extremely low or extremely high earnings (below the 5th or above the 99th percentile).

2.5.3 Main independent variable: physical attractiveness

Critically for the present study, the pairfam dataset includes a measure of physical attractiveness which we use as our primary independent variable. In wave 1, the interviewer was asked, "How attractive do you find the respondent?" The interviewer then reported the respondent's attractiveness on a 7-point scale, from 1 (very unattractive) to 7 (very attractive), with no additional instructions provided. Despite being used widely (Doorley & Sierminska, 2015; Gehrsitz, 2014; Harper, 2000), such a measurement is not ideal as it relies on one person's subjective opinion rather than the mean of many such opinions (often called the "truth of consensus" approach (Patzer, 1985, p. 17)). This shortcoming represents a limitation of our work. To account for the fact that different interviewers may rate faces differently, we include fixed effects or random effects for each interviewer in our models. Additionally, in some model specifications we exclude 513 interviews from 22 interviewers who have very little variation in their attractiveness ratings ($\sigma < 0.4$, σ in full sample = 1.4). In some analyses, we create a binary variable in which we divide respondents into those rated 7 out of 7 (very attractive) and all others (those rated 1-6). We feel this is an appropriate division because our measure is skewed toward high levels of attractiveness (see descriptive results in Appendix A, and Figure A.1), and because we expect attractiveness to be most salient for the highly attractive.

2.5.4 Covariates

To evaluate the intersectional effects of attractiveness, we also study interactions of physical attractiveness and ethnicity. Data on ethnicity is constructed by pairfam and based on information about respondents' countries of birth, as well as those of their parents. Following the commonly used German definition of "migration

background," respondents are classified as being an "immigrant" (or a direct descendant) if the respondent was born outside of Germany or if one or both of their parents were born outside Germany. Pairfam divides respondents into five groups: German natives, Ethnic German immigrants ("Aussiedler", referring to people of German ancestry who lived mostly in Eastern Europe and who returned to Germany in the second half of 20th century, as well as their descendants), multiethnic (so-called "Half-German" in pairfam, i.e., one parent is a native German and the other parent is not), Turkish background, and other non-German background. We use these classifications to study how attractiveness impacts specific ethnic groups, as described in hypotheses 2A and 2B.

We attempt to control for those factors that impact both attractiveness and earnings. Broadly, these covariates relate to respondents' socio-economic status (SES), gender, health, personality, and social distance. As social background has long been linked to perceived physical attractiveness (E. M. Hill et al., 1987; Schunck, 2016), we include parental educational attainment and own educational attainment. However, respondent education may also be a mechanism: if attractive students are treated favorably in school, this might in turn lead to better educational outcomes (Langlois et al., 2000). Therefore, we vary the inclusion of respondent education in the multiverse analysis. We note that parental educational attainment is not sufficient to account for all of the ways in which parental background might influence both earnings and perceived physical attractiveness. To measure health, we include a question asking about respondents' health over the past month (on a 5-point scale) as well as their body mass index (BMI). We also include measures of Big Five personality traits constructed from a 13-item personality assessment. It is not clear whether the respondent's personality operates as a mechanism or a confounder, as it may have been shaped by the effects of physical attractiveness but may also contribute to interviewers' perceptions of their attractiveness (Fletcher, 2013). As such, we test models with and without personality variables. As a measure of social distance beyond the respondent's immigrant status, we also include measures of respondent's nationality (citizenship) and religiosity as well as their German language proficiency reported by the interviewer in wave 1 (average of 4-point scales of speaking and understanding).

Finally, we acknowledge that we cannot directly control for a final potential confounder: cognitive capacity. Cognitive capacity has an obvious connection to work productivity, but may also be linked to physical attractiveness (Kanazawa, 2011). Pairfam data does not include any direct measure of cognitive ability, so instead we include age, height, and interview duration as proxy variables. Previous research has found that taller people tend to have higher cognitive capacity due to factors like childhood health and nutrition (Lundborg et al., 2014; Schick & Steckel, 2015), and height has also been linked to perceptions of attractiveness, especially for men (Jæger, 2011). As our measure of height, we use the largest measure reported over all waves. Cognitive capacity has also been associated with interview duration

(Loosveldt & Beullens, 2013), although we expect that its effect will be nonlinear. People with lower cognitive capacity may speed through the interview by using heuristics or they may alternatively take longer to understand and respond to questions. To account for this, we use the squared interview duration in our models. This adjustment strategy is suboptimal as it requires the assumption that there is no additional causal path from cognitive capacity to earnings, i.e., that the measurement error introduced by using proxy variables is independent and nondifferential (Hernán & Cole, 2009). Therefore, we vary the inclusion in the multiverse analysis. Descriptive statistics on the covariates for our preferred model specifications can be found in Appendix A (Tables A.1 & A.2).

2.5.5 Models

We restrict the model space to linear regression models that are adjusted to fit the clustered structure of our data (Wooldridge, 2015). In pairfam, each observation is nested within respondents and interviewers, creating a three-level structure. A general linear model can be characterized by

$$y_{ijt} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_{ij} + \beta_2 z_{ij} * x_{ijt} + \beta_4 t + \gamma H' + u_i + v_j + \varepsilon_{ijt}$$

with i indicating respondent, j indicating interviewer, and t indicating period. y_{ijt} is the outcome (the natural logarithm of gross hourly earnings), x_{ij} is the interviewer rating of respondent attractiveness and z_{ij} is ethnicity. H' is a vector of control variables. The model contains a respondent effect (u_i), an interviewer effect (v_j), and the idiosyncratic error (ε_{ijt}). Unbiased estimates would require strict exogeneity. Since our interest lies in the interaction between attractiveness and ethnicity, we estimate and present the marginal effects $\delta y_{ijt}/\delta x_{ij}$.

  We estimate three types of models. We first estimate random effects models with random intercepts for respondents and fixed effects for survey waves, adjusting standard errors for clustering within respondents (Wooldridge, 2015). Second, we estimate models with random intercept for interviewers, wave fixed effects, and cluster robust standard errors that adjust for clustering in interviewers. Third, we estimate models with interviewer and wave fixed effects, with standard errors adjusted for two-way clustering in respondents and interviewers (Correia et al., 2020). We cannot include respondent fixed effects as our measurement of physical attractiveness is time-invariant (measured in wave 1). All models are estimated separately for female and male respondents.

  To address a potential selection bias in female labor force participation, in particular for immigrant women, we estimated the above models for female respondents using stabilized inverse probability weights (IPW) for labor force participation (Hernan & Robins, 2020). To do so, we predicted the conditional probability of each woman's participation in the labor force for every period and weighted each

observation by the inverse (reciprocal) of the conditional probability of labor force participation, with the stabilizing factor being the unconditional labor force participation probability (Hernan & Robins, 2020). All models are additionally weighted by the calibrated design weights to address the different cohort inclusion probabilities, to adjust the data to the target population, and to control for baseline survey participation as well as panel attrition (Brüderl et al., 2021).

Table 1 shows the data-analytic decisions described above, our preferred specifications (in italics), as well as the alternatives. Our preferred specification uses a fixed-effects model with interviewer fixed effects and standard errors adjusted for two-way clustering, applied to all adult (18+) respondents in the main sample. Observations from interviewers with no variation in ratings of respondent attractiveness are excluded. With regard to covariates, we use the natural logarithm of gross hourly earnings and a binary measure of attractiveness, and we include respondent education, the measure of personality, and a proxy for cognitive ability. Results from this model are highlighted in the specification curves shown in Figures 2.1-2.6 and A.2-A.5.

2.6   Results

We present the findings of our multiverse analysis as specification curves (Simonsohn et al., 2020) in Figures 2.1-2.6, as well as Figures A.2-A.5 in Appendix A, and we summarize these results in Tables 2.2 and 2.3. While the bottom panel of the figures depicts information about the specifications, the top panel reflects estimated effect sizes. The specifications are sorted by effect size in ascending order. In total, we estimated 2,592 specifications: 1,728 for female respondents and 864 for male respondents, as female respondents have one additional dimension (IPW).

In general, we find strong evidence that native Germans receive a beauty premium, while the results for immigrants and their descendants are ambiguous. For native German men, we find robust evidence for a positive association between attractiveness and earnings. The estimate of our preferred specification is 0.06 and it is statistically significant at the 5% level (Figure 2.1) – which corresponds to approximately 6% higher earnings for attractive German men compared to their less attractive counterparts. The median effect size across all specifications is 0.04 (median p-value 0.00); 89% of the 864 specifications return statistically significant estimates and the sign stability is 100% (Table 2.2).

Considering ethnicity, there is weak evidence for an association between physical attractiveness and hourly earnings in Ethnic German immigrant men. The effect size of our preferred specification is 0.11, which is statistically significant at the 5% level (Figure 2.3), meaning that very attractive Ethnic German men have 11% higher hourly earnings compared to less attractive Ethnic German men. The median effect size of all 864 specifications is 0.07 (median p-value: 0.04). 58% of the specifications return estimates that are statistically significant and the sign stability is 100% (Table
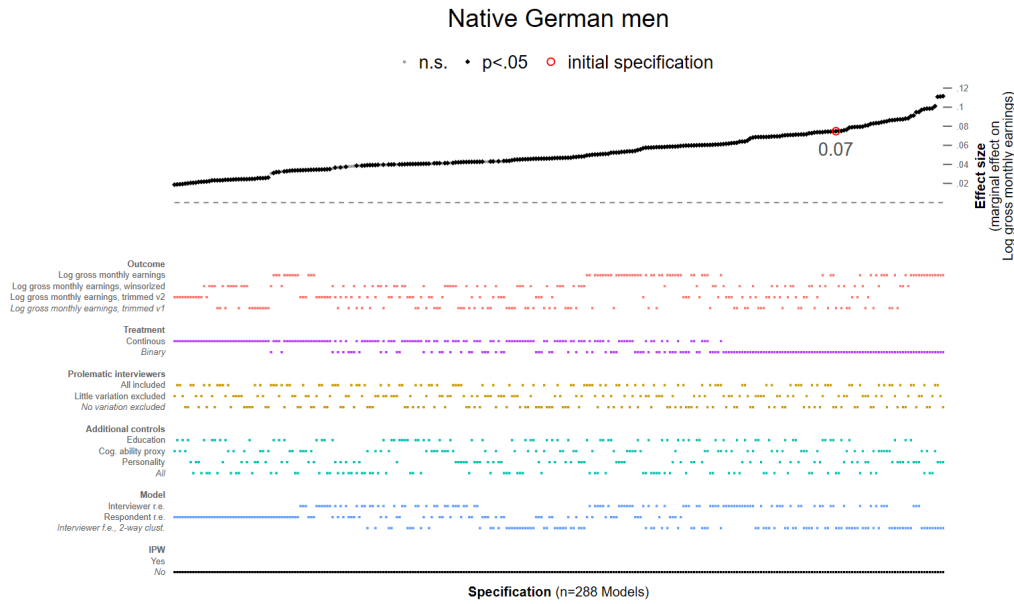
FIGURE 2.1: Specification curve: native German men. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.
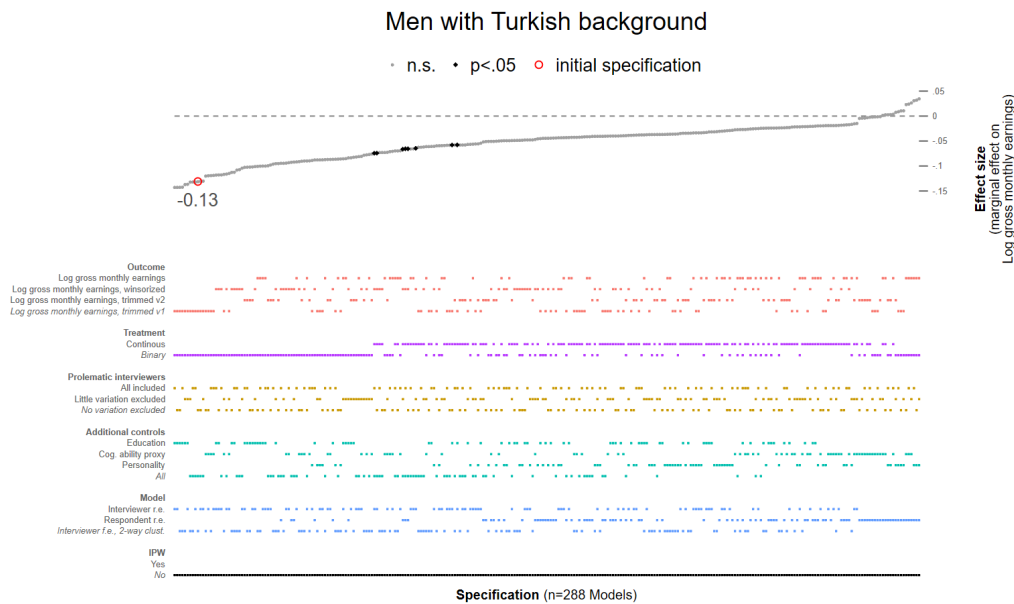
FIGURE 2.2: Specification curve: men with Turkish migration background. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.
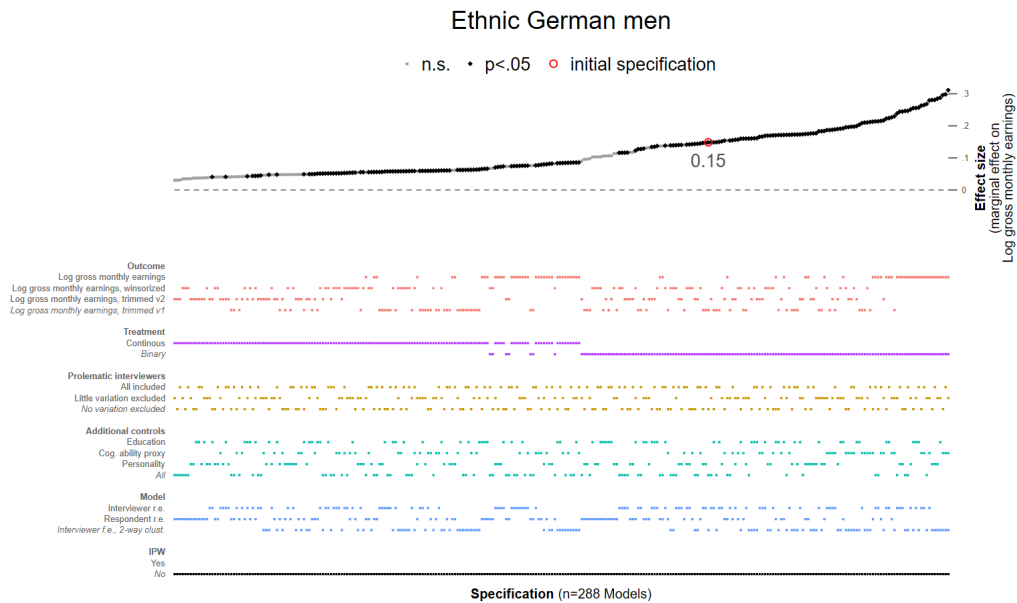
Ethnic German men

FIGURE 2.3: Specification curve: Ethnic German immigrant men. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.

Native German women
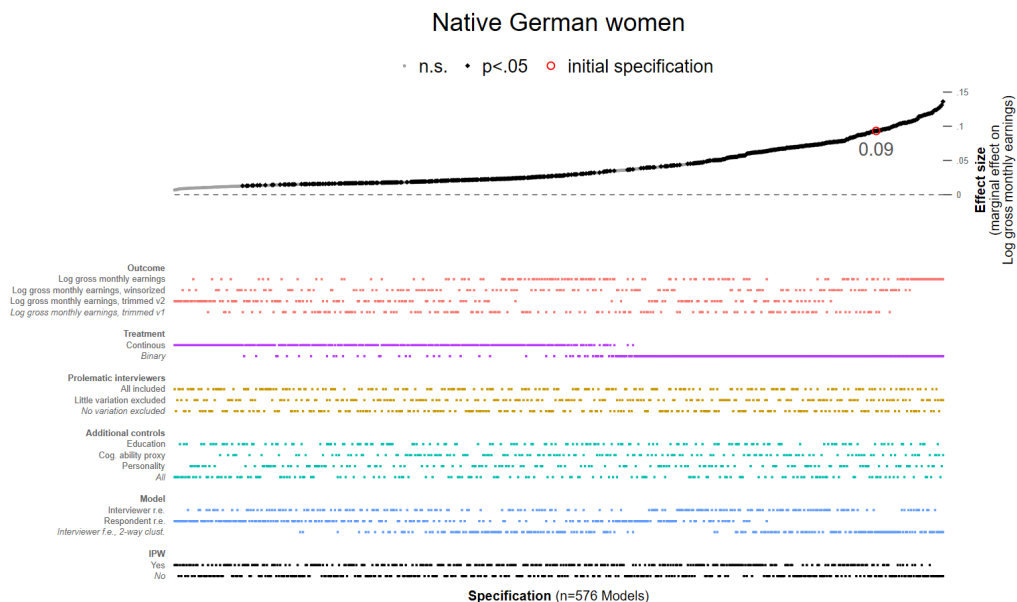
FIGURE 2.4: Specification curve: native German women. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.
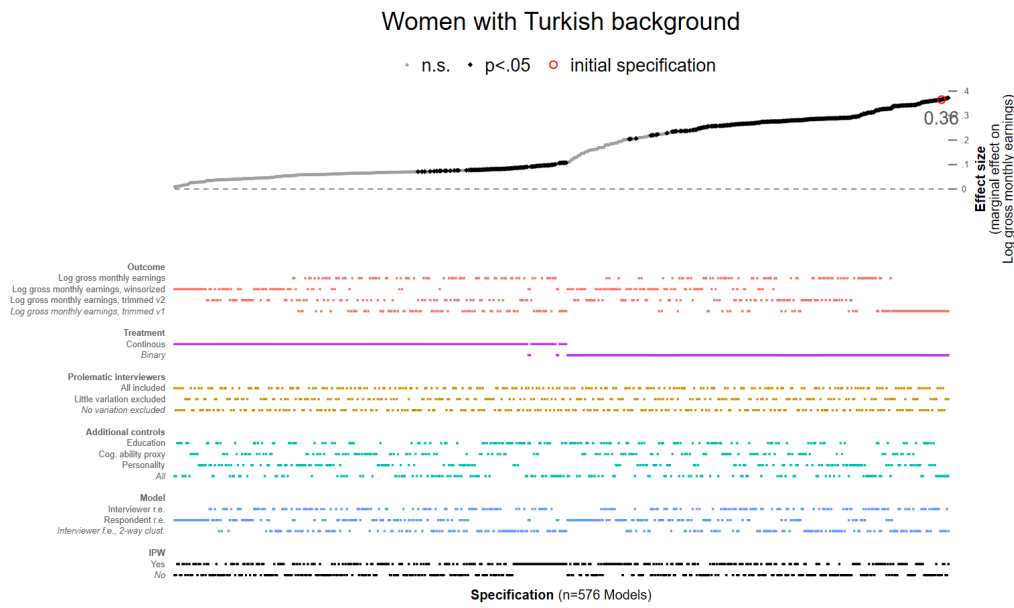
FIGURE 2.5: Specification curve: women with Turkish migration background. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.
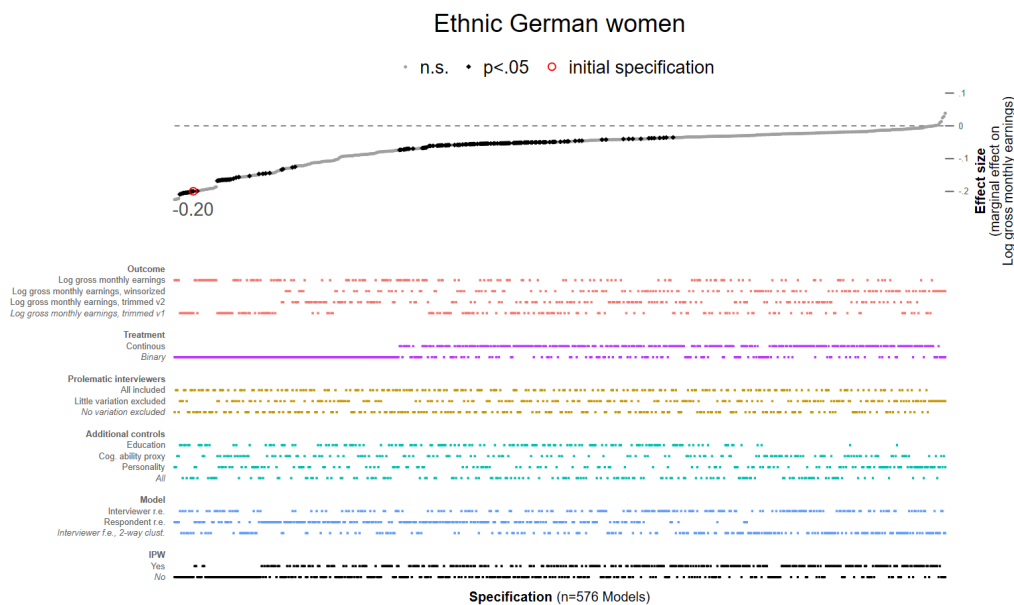
FIGURE 2.6: Specification curve: Ethnic German immigrant women. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.

2.2). When interpreting this finding, we have to take into account that the number of cases for the ethnic minority groups is considerably smaller than for native Germans. For instance, in our preferred specification, we only have 339 observations for Ethnic Germans (Table A.2). For the other ethnicities — immigrants of Turkish origin, multiethnic immigrants, and immigrants of other non-German origin — the analysis robustly returns null results with the vast majority of effect estimates being statistically insignificant (Table 2.2, Figures 2.2, A.2, and A.3).

For native German women, we also find relatively robust evidence for a positive association between physical attractiveness and hourly earnings. Our preferred specification returns an effect size of 0.08 and it is statistically significant at the 5% level (Figure 2.4). The median effect size is 0.03 (median p-value=0.00) and 86% of the estimates are statistically significant (Table 2.3). For other ethnic groups, the results are quite different: at 0.08 (median p-value: 0.09, see Table 2.3), the median effect size of the association of physical attractiveness and earnings for Turkish women is large; even larger is the estimate of our preferred specification at 0.2 although it is statistically insignificant (Figure 2.5). There is some evidence for a beauty premium for Turkish women; however, it is essential to note that the statistical uncertainty is high as only 34% of the estimates return statistically significant results (Table 2.3). The effect sizes for the other ethnic groups are small and the majority of the specifications return statistically insignificant results (Table 2.3, Figures 2.6, A.4, and A.5).

Overall, we find evidence for hypothesis 1B: native Germans receive a beauty premium. There seems to be mixed evidence in favor of hypothesis 2A, at least for women: Turkish women receive the largest benefit from attractiveness; however, only 34% of the estimates return statistically significant results. The aforementioned finding also provides some evidence for hypothesis 2A, as the effect sizes for Turkish women are larger than those for male immigrants and all natives. However, the statistical uncertainty is too high in order to claim that hypothesis 2A is confirmed. For men, on the other hand, we find some weak evidence for hypothesis 2B, as half of the specifications return statistically significant results, indicating that very attractive Ethnic German men may earn more than their less attractive peers.

2.7   Discussion

In this study, we used nationally representative longitudinal data from Germany to examine whether the size of the beauty premium in earnings varies across groups defined by immigrant status. Our multiverse analysis finds strong evidence that native Germans receive a beauty premium. Highly attractive natives earn 3-4% more than their less attractive counterparts. However, we find mixed evidence for a beauty premium among immigrants.

The data provides no clear evidence for many of our hypotheses about the relative size of the beauty premium in various ethnic groups. While we find some evidence for a premium in two groups, only 34% of models for Turkish women and

58% of models for Ethnic German immigrant men show statistically significant results. These findings contradict our hypotheses in that we expected to find an interaction with cultural distance, i.e., that groups with either the lowest or highest cultural similarity would receive the largest premium. Instead, we find some evidence for premia in one group with relatively high similarity (Ethnic German men) and one group with lower similarity (Turkish women). This pattern could suggest diverging mechanisms for men and women, such that resource substitution better explains patterns for women while human capital better explains those for men, or that cultural distance is simply not a relevant factor in determining the size of the beauty premium. It may also relate to gendered ethnic stereotypes in the German context: Muslim men are often seen as aggressive and threatening, which may disadvantage them relative to Muslim women as well as the other groups studied here (Di Stasio & Larsen, 2020).

Our results differ from previous research on the beauty premium examining the role of race and gender in the U.S., which found some empirical support for the resource substitution hypothesis, namely that Black women received the largest beauty premium (Monk et al., 2021). These diverging findings may reflect differences between the American and German labor markets. Employees in Germany may benefit from a more regulated labor market, where wages are often set by collective bargaining agreements and thus cannot be as easily affected by individual factors like perceived attractiveness (Antonczyk et al., 2010; Melzer et al., 2018). On the other hand, this pattern may also be explained by differences in the nature of the two ethnoracial group boundaries. While migrants in Germany face significant discrimination and barriers to full participation in society, their situation is quite distinct from that of Black Americans, who have been excluded from accruing social and material resources for many generations. Migrants may also face different barriers to labor force participation than native racial minorities. While the vast majority of Black Americans are native English speakers with U.S. education, at least first-generation migrants to Germany are unlikely to be fluent in German and may face difficulties having their foreign qualifications recognized (Chiswick & Miller, 2009). This may lead to a wage ceiling for migrants that physical attractiveness cannot overcome.

While our findings are mixed, they seem to provide more support for the human capital perspective than the resource substitution perspective in that we find stronger evidence for a beauty premium among native men and women. This suggests that the value of physical attractiveness may indeed be culturally specific, and that conforming to local beauty standards is critical in order to receive an increase in earnings. These standards may place value on traits and/or grooming practices that are seen as highly typical of the local population and devalue others which are perceived as foreign. In this way, physical attractiveness may function much like foreign educational credentials or work experience, forms of capital that typically lose value when transferred across national boundaries.

It is also important to note that our dataset comes with some significant limitations. First, our measure of attractiveness relies on a single observation by a single interviewer rather than the "truth of consensus" approach more commonly used in the attractiveness literature (Patzer, 1985, p. 17). These interviewers' impressions could be confounded by unobserved factors, especially status cues (Schunck, 2016). These cues may also affect men and women differently, further confounding our results regarding gender differences in the beauty premium. We also lack information on skin color, leaving us unable to evaluate whether race has a significant impact on the beauty premium in Germany as it does in the U.S. Second, the pairfam data groups immigrant respondents into fairly heterogeneous groups. Groups like Ethnic German immigrants, "multiethnic," and "other" may conceal a great deal of variation in terms of cultural similarity, which may be highly relevant to perceptions of physical attractiveness and earnings.

Despite these limitations, this paper makes important contributions to the literature on the beauty premium and labor market inequalities. Our analysis is the first to examine variation in the beauty premium with respect to immigrant status, a highly salient boundary in European societies. While we find that intersections of ethnicity and gender do not seem to structure the beauty premium in the German context, we uncover strong evidence that native Germans are privileged in terms of earnings. However, most immigrants do not seem to receive such a premium. These results suggest that perceived physical attractiveness may be an important mechanism driving persistent immigrant-native wage gaps in Europe.

TABLE 2.1: Overview of potential decision points and alternative decisions.

Dimension	Specifications		Count
Coding	Outcome	*1 - ln gross hourly earnings trimmed* 2 - ln gross hourly earnings winsorized 3 - ln gross hourly earnings 4 - ln net hourly earnings trimmed 5 - ln net hourly earnings winsorized 6 - ln net hourly earnings	6
	Treatment	*1 - binary measure of attractiveness* 2 - continuous measure of attractiveness	2
Covariates	Set of confounders	*1 - age, cohort, parental education, migration status, self-assessed health, BMI, height (cm), period dummies, sample (cohort), interview mode*	1
	Proxy for cognitive ability	*1 - proxy measure included* 2 - proxy measure excluded	2
	Personality	*1 - included* 2 - excluded	2
	Education	*1 - included* 2 - excluded	2
Exclusion criteria	Age restriction on sample	1 - exclude respondents <18 years 2 - include all respondents	2
	Interviewers	*1 - exclude interviewers with no variation in attractiveness ratings* 2 - exclude interviewers with very little variation in attractiveness ratings 3 - include all	3
Missing data handling		1 - listwise deletion	1
Model	Type of regression model	*1 - interviewer fixed effects, s.e. adjusted for two-way clustering* 2 - respondent random effects, cluster robust s.e. 3 - interviewer random effects, cluster robust s.e.	3
Weighting	Application of design weights	*1 - calibrated design weights applied*	1
	Inverse probability weighting	*1 - inverse probability weighting (likelihood to participate in labor force, for women only)* 2 - no inverse probability weighting	2

Note: Initial / preferred decisions are italicized. Note that decisions without alternatives are also listed. Multiple imputation was not feasible for our analysis. Estimating thousands of models for each imputed data set would take several months given the computational resources at our disposal.

TABLE 2.2: Summary of multiverse results for men, by ethnicity.

	Median b	Median p-value	Significant (proportion)	Sign stability (proportion)
Native German	0.04	0.00	0.89	1.00
Ethnic-German (Aussiedler)	0.07	0.04	0.58	1.00
Multiethnic ("Half-German")	-0.04	0.30	0.02	0.98
Turkish origin	-0.02	0.59	0.02	0.65
Other origin	0.04	0.39	0.08	0.84
N (Specifications)	864			

TABLE 2.3: Summary of multiverse results for women, by ethnicity.

	Median b	Median p-value	Significant (proportion)	Sign stability (proportion)
Native German	0.03	0.00	0.86	1.00
Ethnic-German (Aussiedler)	-0.03	0.23	0.13	0.91
Multiethnic ("Half-German")	0.00	0.59	0.01	0.57
Turkish origin	0.08	0.09	0.34	1.00
Other origin	-0.02	0.34	0.13	0.75
N (Specifications)	1728			

Chapter 3

Pretty Qualified? The Role of Attractiveness, Information, and Stereotypes in Ethnic Hiring Discrimination

Abstract

Ethnic discrimination remains a persistent challenge in European labor markets, limiting employment opportunities for immigrants and their descendants. While extensive research has documented ethnic penalties in hiring, less is known about how physical attractiveness interacts with ethnicity to influence labor market outcomes. Prior studies indicate that attractive applicants often receive preferential treatment, but it is unclear whether this "beauty premium" applies equally across ethnic groups. This study fills this gap through a large-scale field experiment in the German labor market, a setting in which applicants are expected to include photos on their résumés. Exploiting this custom, nearly 4,000 fictitious job applications were submitted, varying applicants' ethnic background, gender, attractiveness, and information about their prior performance. Our results show evidence of an ethnic hierarchy: Applicants of Turkish descent face significant discrimination, as do men with a Greek background, while Danish-heritage applicants face no discrimination relative to the ethnic majority. While there is a modest beauty premium in overall callback rates, this premium does not close the German-Turkish ethnic gap. This study contributes to the literature on labor market discrimination and status characteristics, illustrating how seemingly advantageous traits can reinforce ethnic stratification rather than mitigating it.

3.1   Introduction

Immigrants and their descendants face discrimination and persistent inequities in European societies. This discrimination takes a variety of forms, from "everyday" instances of discrimination such as physically avoiding (Zhang et al., 2022) or not offering help to minority group members (Choi et al., 2019), to more systemic forms such as exclusion from the rental housing market (Auspurg et al., 2019). These patterns of differential treatment cannot be entirely explained by linguistic or cultural differences, as even the children of immigrants still report significant discrimination despite high levels of integration into the host society (Heath et al., 2008; van Tubergen, 2025).

A robust literature also documents significant ethnic discrimination on the labor market (Lippens et al., 2023; Zschirnt & Ruedin, 2016). This discrimination is ubiquitous and persistent over years, with audit studies finding that employers have a preference for native majority group applicants even when compared to equally qualified ethnic minority applicants (Kaas & Manger, 2012; Koopmans et al., 2019; Quillian & Midtbøen, 2021; Thijssen, Lancee, et al., 2021). The extent of this discrimination may however depend on the constellation of gendered ethnic and religious stereotypes in a particular context. For example, in Europe, while some studies find that men with Middle Eastern or North African heritage are the most disadvantaged (Bursell, 2014; Dahl & Krog, 2018; Di Stasio & Larsen, 2020), other studies show that women who wear a veil face significant discrimination (Fernández-Reino et al., 2023; Weichselbaumer, 2020).

When making hiring decisions, ethnicity is only one of many attributes of an applicant that employers might consider alongside formal qualifications. Another important aspect of an applicant's appearance is physical attractiveness, a factor which has been less frequently studied but which nonetheless has a formidable effect on interpersonal perception. People have a tendency to ascribe more positive attributes to attractive people, assuming they are both warmer and more competent than others (Dion et al., 1972; Eagly et al., 1991; L. A. Jackson et al., 1995). It is then hardly surprising that attractive people are also favored in the labor market, receiving higher callback rates for job applications, higher pay, and a greater likelihood of promotion (Hosoda et al., 2003; Nault et al., 2020). As with ethnic discrimination, patterns of "lookist" discrimination are also gendered: contrary to the commonly held assumption that attractiveness matters more for women, men receive a more consistent benefit from their looks (Kühn & Wolbring, 2024; Kukkonen et al., 2024).

While certain traits, like facial typicality, are commonly linked to physical attractiveness across contexts (Todorov et al., 2015), perceptions of attractiveness within a given culture typically favor the appearance of locally dominant social groups (Monk et al., 2021). One of the most prominent examples of this is colorism, a widespread preference for lighter skin that is often tied to Eurocentric standards of beauty (Dixon & Telles, 2017). Similarly, members of visible ethnic minority groups

may be less likely to possess other traits that are locally valued, such as specific hair or eye colors, making them less likely to be perceived as highly attractive in their country of residence (Ranzini & Rosenbaum, 2020). Thus, perceived physical attractiveness in a given setting may be unequally distributed across ethnic groups.

In this study, we examine how ethnicity and physical attractiveness interact with respect to hiring outcomes. While previous research has established that both of these highly visible characteristics contribute to unequal treatment on the labor market, less is known about how they interact. Beyond the uneven distribution of attractiveness across ethnoracial groups, recent work also suggests that the benefits of physical attractiveness are not equally distributed between racial groups in the U.S. (Kunst et al., 2023; Monk et al., 2021). This raises the question of whether similar patterns exist for ethnic groups, a group boundary that is particularly salient in the European context. Understanding of the intersection of ethnicity and physical attractiveness is important to gain a deeper understanding of complex labor market dynamics.

As both ethnicity and physical attractiveness have been identified as diffuse status characteristics (Berger & Fişek, 2006; Webster & Driskell, 1983), this research may also contribute to the development of status characteristics theory. By examining the effects of these two status characteristics, we add to an ongoing discussion about how status characteristics combine to shape patterns of disadvantage (Marquis et al., 2024; Pedulla, 2018). This question is also highly relevant to the growing literature on intersectionality: We adopt an inherently intercategorical approach that considers the effects of attractiveness, gender, and ethnicity on inequalities both within and between groups (McCall, 2005).

We study this effect through a field experiment in the German labor market, taking advantage of the common practice of including a photo on one's résumé. This allows us to send nearly 4,000 applications to jobs in eight occupations, varying the fictional applicant's physical appearance while keeping other characteristics constant. Our approach avoids some of the typical limitations of survey experiments, such as social desirability bias, which can lead to underestimations of discrimination (Forster & Neugebauer, 2024). By randomly assigning treatments and testing in a real-world setting, field experiments like ours remain the gold standard to causally identify differential treatment in the labor market.

The results of our experiment demonstrate the importance of considering the intersection of multiple characteristics when studying labor market inequalities. Our findings echo prior research that finds an advantage for more attractive applicants, and that shows clear evidence of an ethnic hierarchy in hiring practices. However, we find that these patterns of ethnic discrimination are also gendered: Ethnic minority men face much stronger penalties than do ethnic minority women, with the exception of women wearing a hijab. Finally, we find that despite our expectations, attractiveness affects all candidates roughly equally, but may offer a larger benefit to

ethnic majority women. Thus, attractiveness does not seem to help close the persistent ethnic gaps in hiring.

3.2 Appearance on the labor market

Physical appearance is a central but understudied factor in explaining hiring discrimination. People create remarkably stable social evaluations of others based on even brief exposure to a face (Todorov et al., 2015; Zebrowitz & Montepare, 2008). These evaluations are highly relevant in hiring situations, where recruiters must rely on limited information about an applicant to quickly assess who might meet the requirements of the job. While appearance is perhaps more relevant in settings where photos are included in application materials, such as many European and Asian countries, it is also increasingly relevant worldwide due to the widespread use of social media for screening candidates (Acquisti & Fong, 2020; Baert, 2018).

Appearance is used as a heuristic because it can convey a wealth of information about a person. This information allows us to place the person into various social categories, each of which may be linked to stereotypes that allow us to quickly (if not necessarily accurately) judge a person on various dimensions (Todorov et al., 2015). Appearance conveys information about structural and identity-based factors like ethnicity and gender that are frequently studied in sociological research. Both ethnicity and gender have been linked to substantial inequalities in labor market outcomes in Europe and beyond (Lippens et al., 2023; Zschirnt & Ruedin, 2016). Similarly, physical markers of difference such as phenotype and skin color may also influence a person's treatment on the labor market (Dias, 2020; Polavieja et al., 2023).

Appearance also necessarily conveys information about physical attractiveness, a dimension which is less frequently studied despite its powerful influence on social evaluations (Hamermesh, 2011). This research gap is surprising given that attractiveness is likely to intersect with these other social characteristics in shaping perceptions of status and employability. Addressing this gap, this study examines how appearance may exacerbate or mitigate ethnic and gender penalties on the labor market. Drawing on theories of status characteristics and stereotyping, we examine how attractiveness intersects with ethnicity and gender in the context of hiring.

3.2.1 Appearance as status

Status Characteristics Theory (SCT) is a useful theoretical framework to explain how we use information about a person's traits to form expectations about their behavior and performance. SCT conceptualizes attributes like physical attractiveness, gender, and ethnicity as "diffuse status characteristics," or socially salient characteristics with clearly defined states that are differentially evaluated in terms of social status (Berger & Fişek, 2006; Webster & Driskell, 1983). People who have higher-status characteristics are assumed to be more competent and more worthy of resources than those with lower-status characteristics. These cultural stereotypes may

influence selection processes and thus reinforce status-based inequalities (Correll & Ridgeway, 2006). While these status characteristics may have no connection to actual competence, we form consensual beliefs about the relative status of various groups, such as the widespread belief that men are higher in status than women (C. Ridgeway, 1991). Similarly, highly attractive people and members of the ethnic and religious majority group are thought to be higher in status than less attractive people and members of minority groups, respectively (Berger & Fişek, 2006; Webster & Driskell, 1983).

Initially, status characteristics theory focused on predicting the effects of single characteristics, but over time scholars have expanded the theory to account for the effects of multiple characteristics. This more clearly reflects real-world conditions, in which it is rare to have information about only one trait. Even a photo or a name might convey information about multiple traits, including but not limited to gender, ethnicity, age, and social class, all of which might be relevant to status expectations. Status characteristics theory supposes that when a person is evaluated, they are evaluated on the basis of all status information simultaneously (Webster & Driskell, 1978). Generally, this assumes that the effects of status characteristics are additive in nature: People with the greatest number of high-status characteristics will be seen as highest in status and will thus receive the most positive outcomes, and vice versa for those with the greatest number of low-status characteristics. However, each additional characteristic has a smaller marginal impact, so that those with many high-status characteristics should benefit less from adding one more (Berger et al., 1980).

Research on status characteristics also draws a distinction between "diffuse status characteristics" like attractiveness and ethnicity, which influence expectations across all types of situations, and "specific status characteristics" like language proficiency or arithmetic skill, which influence expectations only with respect to certain tasks. Specific status characteristics are assumed to have a stronger impact on expectations of competence in relevant tasks (Berger et al., 1980). This is important to consider in the labor market context, as applicants necessarily include information about their competences and past performance as well as information about their identity. From this theoretical perspective, we would expect performance to outweigh identity, although some research has also found that such information has little effect on ethnic discrimination (Thijssen, Coenders, & Lancee, 2021a).

This additive conception of status would suggest that high-status characteristics like physical attractiveness might counteract some of the detrimental effects of low-status characteristics, such as belonging to a minority ethnic group. This does not mean, however, that one's status is determined simply by the sum of "positive" and "negative" characteristics, as recent work suggests that the effect of each characteristic varies in size. The strength or weakness of a given characteristic depends on its congruence with other status characteristics and with the context in which these characteristics are judged. In a study about race and unemployment in the U.S.,

Pedulla (2018) finds that periods of unemployment harm Black applicants less than White applicants on the labor market. Pedulla attributes this effect, termed "muted congruence," to the fact that stereotypes about the unemployed and about Black applicants strongly align, and thus this added status information has smaller marginal impact, as suggested by earlier work on status organizing principles (Berger et al., 1980). More recently, Marquis et al. (2024) complicate the picture even further by suggesting that applicant-job fit depends not only on the status of the applicant but also how well it matches the status of the job. They find that applicants with mixed-status characteristics (some high, some low) are penalized in both high-status and low-status positions as employers cannot easily assess their fit to the position in question. Together, these studies illustrate the complex mechanisms underlying the combination of status characteristics, problematizing the notion that combining effects is a matter of simple addition.

3.2.2 Appearance and intersectionality

Research in the tradition of intersectionality theory also rejects this additive perspective in favor of a "complex" approach (Heiserman, 2023; McCall, 2005). Instead of considering the effects of a single characteristic at a time, intersectional scholars propose that interpersonal evaluation depends on the entire constellation of a person's characteristics (Crenshaw, 1989; Kang & Bodenhausen, 2015; McCall, 2005). For example, social perceptions of Black women are not necessarily the same as stereotypes of Black people and stereotypes of women combined (Ghavami & Peplau, 2013). Instead, these categories are considered simultaneously, producing unique gendered racial stereotypes (Browne & Misra, 2003; C. L. Ridgeway & Kricheli-Katz, 2013). This conception of the social world demands new analytical approaches, such as the "intercategorical" approach which examines the relation of various social categories both within and across categories of interest (McCall, 2005).

Although such research insists on the intertwining of social categories, some categories seem to be more important than others in forming interpersonal perceptions. The categories most often mentioned as "primary" categories are age, race or ethnicity, and gender (Heiserman, 2023; Kang & Bodenhausen, 2015). These traits are seen as particularly likely to structure the contents of stereotypes about other "secondary" traits like class, which are perceived more slowly and with less certainty (C. L. Ridgeway & Kricheli-Katz, 2013). Although attractiveness is quickly assessed, we process information about gender first, making it likely that attractiveness will also act as a secondary category (Carbon et al., 2018).

Unlike status characteristics theory, this intercategorical approach allows for the possibility that characteristics may have different effects across social groups. Status characteristics theory would suggest that as a high-status characteristic, attractiveness always has a positive effect, even if the size of this effect varies. In an intercategorical analysis, the sign of this effect may change direction for members of certain groups depending on the unique stereotypes of that group.

This is a particularly important possibility to consider with respect to physical attractiveness, as previous results have indicated that gender has substantial effects on beauty-related rewards. Specifically, some work on the so-called "beauty is beastly" effect finds that women are occasionally penalized for attractiveness on the labor market, while men are consistently rewarded for it (Heilman & Saruwatari, 1979; S. K. Johnson et al., 2010; Kukkonen et al., 2024). On top of these gendered effects, the effects of attractiveness may also be stratified by race and ethnicity: Research on wage setting finds that Black women benefit the most from attractiveness in the U.S. (Kunst et al., 2023; Monk et al., 2021), while in Germany, ethnic majority group members benefit more than those from minority groups (Hellyer et al., 2023). It has been suggested that the value of attractiveness relates to the perceived quantity of attractiveness within each group: If a group is stereotyped as unattractive, this means that beauty is a scarce resource within the group, making it all the more valuable for the rare group member that has it (Monk et al., 2021). Because perceptions of attractiveness are linked to power relations in a given society, beauty standards tend to be modeled after higher-status groups, such as the local ethnic majority (Mears, 2014). This means that the beauty premium is inherently relational and intersectional: Its value depends on one's group membership and the position of this group relative to others.

3.2.3 Appearance stereotypes

This conception highlights the critical role of stereotyping in creating and perpetuating appearance-based inequalities. The commonly held stereotype that "what is beautiful is good" is one mechanism through which the beauty premium operates (Dion et al., 1972). In terms of the Stereotype Content Model (Fiske et al., 2002), highly attractive people are thought to be both warmer and more competent than their less attractive peers (Eagly et al., 1991; Langlois et al., 2000; Nault et al., 2020). This perception may affect how others act toward highly attractive people, as those perceived to be warm and competent are more often met with facilitating behaviors like collaborating and offering help (Cuddy et al., 2008).

Stereotypes about warmth and competence also relate to common theoretical explanations for labor market discrimination (Veit et al., 2022). Perceptions that members of minority groups are less warm may drive taste-based discrimination, or an employer's preference for the majority out of simple animus toward minority groups (Becker, 1971). On the other hand, perceptions that minority group members are less competent may drive statistical discrimination, or a preference for the majority due to the belief that they are more productive (Arrow, 1973; Phelps, 1972).

For highly attractive members of stigmatized groups, positive stereotypes about attractiveness may help to offset negative stereotypes of immigrant groups. In Germany, people of Turkish descent (the largest immigrant group in Germany) are thought to be both less warm and less competent than ethnic majority Germans (Asbrock, 2010; Froehlich & Schulte, 2019). Physical attractiveness may then provide a

signal of warmth and competence that causes a highly attractive person of Turkish descent to be seen as an exception to the predominant negative stereotypes of their ethnic group (Pedulla, 2014; C. L. Ridgeway & Kricheli-Katz, 2013). These "exceptions to the rule" may be rewarded on the labor market, serving as "tokens" that diversify the workforce on paper while being seen as relatively similar to the majority group (Monk et al., 2021). Because members of the majority group are already seen as relatively warm and competent, these additional cues may be less significant to their labor market outcomes.

According to theories of statistical discrimination, appearance-based stereotypes, such as those related to physical attractiveness, should exert their strongest influence in low-information situations (Arrow 1973; Phelps 1972). When little else is known about an individual, people tend to rely on visible cues and stereotypes to make judgments. However, as more information becomes available — such as qualifications, skills, and performance records — the influence of appearance-based stereotypes may diminish. While additional information should yield more nuanced and accurate evaluations, experimental evidence for the attenuation of discrimination remains mixed (Lippens et al., 2022; Thijssen, Coenders, & Lancee, 2021a, 2021b).

3.3   Hypotheses

Bringing together literature on status characteristics, intersectionality theory, and stereotyping, we derive several pre-registered hypotheses[1] about the interplay of ethnicity, gender, and physical attractiveness on labor market performance. Following a wealth of evidence, we expect that employers in our field experiment set in the German labor market will prefer applicants belonging to the majority ethnic group over members of ethnic minority groups:

H1: Applicants without a migration background will be preferred to those with a migration background.

Due to pervasive positive stereotypes of attractive people, we also expect that high levels of physical attractiveness will be met with greater success on the labor market:

H2: Highly attractive applicants will be preferred to less attractive applicants.

While we expect that attractiveness will generally be rewarded on the labor market, previous research also suggests that the benefits of beauty are not equally distributed. Gender has frequently been shown to moderate the effects of attractiveness; men seem to receive a fairly consistent beauty premium while attractiveness only benefits women in certain contexts (Kukkonen et al., 2024). In some contexts, attractive women may even face labor market penalties, perhaps owing to perceptions that they are a poor fit for masculine-typed work (Heilman & Saruwatari, 1979;

---

[1]Our experiment is pre-registered on OSF: https://osf.io/p59gj

S. K. Johnson et al., 2010). Because of these varied results on the intersection of gender and attractiveness, we refrain from making a directional hypothesis and simply hypothesize that:

H3: Gender will moderate the effect of physical attractiveness on callback rates.

In comparison to the wealth of research on the interplay of gender and attractiveness, the interaction of ethno-religious identity and attractiveness has rarely been studied. The few studies that address this intersection come to opposing conclusions, perhaps due to contextual differences. While American research finds a particularly large beauty premium for members of stigmatized groups (at least in terms of wages and suitability perceptions) (Kunst et al., 2023; Monk et al., 2021), other research, including a German study, finds a larger beauty premium for higher-status groups (Galarza & Yamada, 2014; Hellyer et al., 2023). Building on research about status characteristics and stereotyping, we expect that the size of the beauty premium in hiring will vary across groups defined by ethnicity and gender. Gendered ethnic stereotypes may combine with attractiveness stereotypes in idiosyncratic ways that affect employers' perceptions of applicants' suitability for the position in question. Attractiveness signals both warmth and competence (Eagly et al., 1991; Langlois et al., 2000; Nault et al., 2020), which may offset negative stereotypes of applicants with migration background, while it may do less to change already positive impressions of ethnic majority applicants. These positive stereotypes may particularly benefit men with a migration background, who seem to face more significant discrimination in Europe (Bursell, 2014; Dahl & Krog, 2018; Di Stasio & Larsen, 2020). Thus, we expect that physical attractiveness may help to close the (gendered) ethnic gap in hiring outcomes:

H4: Applicants with a migration background will receive a larger benefit from attractiveness in terms of callback rates.
H5: Male applicants with a migration background will receive a larger benefit from attractiveness in terms of callback rates.

With this study, we also aim to uncover some of the mechanisms that underlie both ethnic and lookist discrimination on the labor market. As suggested by the theory of statistical discrimination, one such mechanism might be that employers lack necessary information about candidates' productivity, and thus rely on stereotypes as a heuristic (Arrow, 1973; Phelps, 1972). The German labor market is a least likely case for such discrimination due to the extensive and highly standardized documentation included in the typical job application (Zschirnt & Ruedin, 2016). Because of these norms, we are able to systematically vary the amount and quality of this information included in our applications in order to test whether it plays a role in patterns of discrimination. In keeping with statistical explanations of discrimination, we expect that:

H6: Providing positive information about job performance will reduce discrimination on the basis of physical attractiveness.

Employers' preference for attractive candidates may also vary with occupation. Some research suggests that more attractive candidates are more strongly preferred in positions with extensive customer contact (Rooth, 2009; Tews et al., 2009). In these fields, such as sales or food service positions, it may be assumed that attractive workers will be preferred by customers, or that they can be more persuasive, making attractiveness a plausibly productive characteristic. By testing in multiple occupations with varying degrees of customer contact, we can also test whether occupational differences might shape patterns of employer preference:

H7: The effect of attractiveness on callback rates will be stronger in fields with high levels of customer contact.

Finally, we examine the role of religious stereotypes on hiring behavior. Above and beyond the discrimination facing people belonging to ethnic minority groups, there is an additional penalty for candidates that express a strong belief in Islam in Christian-majority European countries (Di Stasio et al., 2021; Valfort, 2020). While religion is not usually included in a résumé, the German custom of including a photo in application materials allows us to include applicants wearing a hijab, a clear signal of adherence to Islam. Unsurprisingly, veiled women experience substantial discrimination in Europe (Fernández-Reino et al., 2023; Weichselbaumer, 2020) and the United States (Ghumman & Ryan, 2013). This discrimination may reflect employers' beliefs that customers will not wish to interact with veiled women, a phenomenon sometimes called "customer discrimination" (Becker, 1971; Combes et al., 2016). Given the level of discrimination reported in previous work, we expect to find similar patterns in our test:

H8: Turkish-origin women wearing a headscarf will receive fewer callbacks than Turkish-origin women without a headscarf.

Discrimination against veiled women may result from simply activating stereotypes about Muslims. However, another yet unexplored possibility is that by covering a woman's hair, veiling lowers perceived attractiveness, triggering both ethnic and lookist discrimination. We vary both attractiveness and veiling in our study in order to explore whether appearance-based discrimination might explain some of these differences in callback rates:

H9: Religiosity (specifically wearing a headscarf) will moderate the effect of physical attractiveness on callback rates.

Dimensions	Levels
Gender	Male / Female [2]
Attractiveness	Low / Medium / High [3]
Ethnicity (Name)	German / Turkish / Greek / Danish [4]
Prototypicality (Phenotype; only candidates with German or Turkish names)	Low prototypicality / High prototypicality [2]
Performance Information	No references / Average / Excellent [3]
Religiosity (only Turkish-heritage female candidates)	No headscarf / Headscarf [2]

TABLE 3.1: Dimensions varied in the correspondence test.

3.4 Methods: correspondence testing

To test these hypotheses, we conduct an unpaired correspondence test on the German labor market. We respond to real job advertisements on the Federal Employment Agency (*Bundesagentur für Arbeit*, or BA) website with applications from fictitious applicants whose characteristics randomly vary in the dimensions listed in Table 3.1. We use an unpaired, randomized design in which each company only receives one application to reduce the likelihood of detection and to minimize the burden on employers (Larsen, 2020; Vuolo et al., 2018). We then track which applications receive positive responses and compare patterns of response across lines of ethnicity, gender, and physical attractiveness.

Germany is a strategic site for our research due to the local norm of including a photo on one's résumé when applying for jobs, allowing us to easily and naturally signal our applicants' facial attractiveness, ethnicity, and religiosity. In addition to the photo, applications use a highly standardized application format including a résumé, cover letter, educational certificates, and (in most cases) reference letters from a previous employer, as is expected in Germany. Each applicant's résumé uses a similar format but contains randomly assigned names and photos displayed prominently at the top of the document. Résumés also include corresponding postal and email addresses and an educational and occupational history that matches the job to which they are applying.

All applicants, regardless of their ethnicity, state in their résumés that they are German citizens who were born in the country and completed their education there. Each applicant states that they completed their occupational training and worked in their first job for 2-3 years. They motivate their application in the cover letter by saying that they would like to move back to the area of the job to be near family. This cover story allows us to apply for jobs even in smaller communities, in which applications from a large city may be seen as implausible. Cover letters, résumés, and educational certificates are identical within each occupation with the exception

of the applicant's name and their level of work performance.

Each application is randomly assigned a fictitious identity belonging to one of four ethnic groups (ethnic majority German, or Danish, Greek, or Turkish migration background) using simple randomization. Each applicant is given a name that reflects their heritage as well as their randomly assigned gender. They are also randomly assigned a photo which also signals ethnicity and gender, as well as a level of physical attractiveness. These ethnicities were chosen for both practical and theoretical reasons. People of Turkish descent represent the largest minority ethnic group in Germany, and ethnic majority Germans are a logical reference group for studying ethnic discrimination. While the Danish and Greek populations in Germany are much smaller, these groups are generally phenotypically similar to Germans and Turks, respectively. This phenotypic similarity allows us to use the same set of pictures for four ethnicities that differ in terms of cultural stereotypes. The inclusion of these ethnicities thus allows us to test whether any effects we find are due to migration background, phenotype, or applicants' position on the perceived ethnic hierarchy in Germany, a factor that has proven influential in previous correspondence tests (Koopmans et al., 2019; Veit & Thijsen, 2021).

We expect that this hierarchy would privilege those of Danish descent, who are seen as similar to Germans and are generally positively stereotyped (Hofmann & Hallsteinsdóttir, 2016). The history of Greek and Turkish migration to Germany is quite similar. Both groups are well-established in the country, as large groups of both Greeks and Turks arrived as guest workers in the decades following World War II. While the German public image of Greeks may have soured somewhat during the Greek debt crisis (Tseligka, 2016), we nonetheless expect a shared Christian heritage and EU membership to create a more positive impression of people with Greek migration heritage when compared to those with Turkish migration background, who are often associated with a variety of negative stereotypes in Germany (Baltes & Rudolph, 2010). Comparative research finds that while Greeks and Turks are viewed to be roughly equally competent in Germany, Germans feel warmer toward Greeks (Froehlich & Schulte, 2019).

Photos were pre-selected from a set of photos collected from three academic databases: the Bogazici Face Database (Saribay et al., 2018), the Chicago Face Database (Ma et al., 2015), and the DeZIM Picture Database: Faces (Veit & Essien, 2022). As shown in Figure 3.1, all photos were edited to wear matching clothing and appear against an identical gray background. After editing, an initial set of 108 photos was rated by an online access sample of 1,125 German residents on physical attractiveness and ethnic typicality (German, Danish, Greek, and Turkish) among other factors like social class, trustworthiness, and gender typicality. The photos were then manually sorted by level of physical attractiveness and ethnic typicality, and a final set of 36 was selected. This set generally includes 2 photos for each combination of ethnicity and physical attractiveness, although some combinations use only one when no suitable matching photo could be found within our dataset. We also include some

FIGURE 3.1: Sample photos used in the correspondence test. Left column: ethnic majority German men. Middle column: Turkish-heritage women. Right column: Turkish-heritage applicants, female in religious condition. Low attractiveness images in top row, high attractiveness in bottom row. Original face imagery from Chicago Face Database (Ma et al., 2015) and DeZIM Picture Database: Faces (Veit & Essien, 2022), edited for consistency.

photos which were rated about equally likely to be German or Turkish, and assign them both German and Turkish names in the test. Following previous research (Ma & Correll, 2011), we label these photos "low prototypicality," as they were not rated as highly prototypical of either German or Turkish faces. These photos provide an even clearer test of the ethnic penalty by keeping facial appearance constant while varying only the ethnically typical name. Finally, for photos used in the Turkish female condition, we also include versions that were edited to add a headscarf, a visual cue of adherence to Islam.[2]

Names were created using lists of popular given names in the mid-to-late 1990s (the birthdate of our applicants) in Germany, Denmark, Greece, and Turkey, as well as lists of common surnames in each country. These names were rated by another online access panel of 800 German residents on ethnic and gender typicality as well as connotations of social class. The final set of 48 names were chosen to ensure that each name provides a clear signal of the intended combination of ethnicity and gender in the German context without being outliers in terms of social class perceptions.

---

[2]We also included some Turkish male profiles with a full religiously styled beard as the closest equivalent treatment for men. However, as evidenced by a post-hoc survey, these beards were not recognized as a signal of religiosity in the German context. Therefore, we do not differentiate between these photos and the other Turkish male profiles.

All applicants are randomly assigned a level of information about work experience (none, average, or high), which corresponds to the presence and quality of employer references included with the application. These reference letters are highly formulaic (as is the custom in Germany) and are written at two levels, one indicating enthusiastic praise of the applicant's work performance ("excellent quality") and another more subdued letter indicating that the applicant meets expectations ("average quality"). In the "no information" condition, the applicant does not include a reference in the application, which is contrary to the norm.

We select job offers from those available on the BA's Jobsuche ("job search") website until we reach our target sample size. We search for each of the eight jobs selected using the suggested search terms (given by the site's autofill function) using a semi-automated approach on each day during the test period. At the collection phase, we exclude temporary positions and positions advertised within 100 km of the city of our applicants' current residence to ensure that the cover story fits.[3] The jobs collected are then manually reviewed to assess whether the job is a good fit for our applicants' profiles in terms of desired education and work experience, and to filter out irrelevant listings, temporary positions, and advertisements posted by a recruiter. Excluding recruiters is necessary as these listings generally do not include information about the name of the employer, making it impossible to assess whether we have contacted the employer before.

We apply to full-time and part-time permanent jobs listed in eight occupations: retail salesperson, receptionist, custodian, secretary, purchasing agent, industrial clerk, sales associate, and IT specialist. These jobs were selected as they are quite common, with a large number of job postings available at any given time, and were not strongly impacted by personnel shortages due to the Covid-19 pandemic. We also used data from the German Microcensus to exclude any occupations in which workers with a migration background were exceedingly rare, to ensure that our applications will fit the typical applicant profile. Finally, these jobs include both male-dominated and female-dominated professions, and those with and without direct customer contact. Gender imbalances in two occupations led us to send only male applications for custodian jobs and only female applications for receptionist jobs in the interest of realism.

To comply with the usual ethical standards for conducting correspondence tests (Zschirnt, 2019), we respond to all employer inquiries with an email response within one business day. Our response thanks the employer and indicates that the applicant has already found another suitable position. If the employer continues to contact the applicant after this point, we do not respond but continue tracking all responses. Responses are manually classified into nine categories using the scheme created for the GEMM study (Lancee et al., 2019, p. 23). Responses are considered positive if they contain a request for an interview, evidence that the applicant has passed an initial

---

[3]Until late March 2024, we also excluded jobs posted by small employers (with fewer than 6 employees) to avoid burdening small businesses. After this time, information about company size was unfortunately no longer publicly available.

selection process, and/or a request for additional information about the candidate. We classify responses as negative if they contain an explicit rejection or if the applicant only receives a confirmation that the application was received with no further responses. Other responses, such as missed calls with no message left, or responses stating that the application is incomplete and cannot be evaluated, are classified as missing. These unclear response types are overwritten by more definitive response types, such that a missed call followed by a pre-invitation is classified as positive and a missed call followed by a rejection is classified as negative.

We sent applications from September 2023 to September 2024, and continued tracking responses until the end of December 2024. During this time, we sent 3,948 applications. Of these, 92 applications included incorrect or missing documents and were removed from the dataset, leaving us with a final analytical sample of 3,856 applications.[4] A more detailed accounting of the number of responses sent in each gender by ethnicity by attractiveness condition is shown in Table B.1. As shown in Table B.2, 1,626 (42.2%) of these applications were met with a positive response, including 1,099 (28.5%) that received explicit requests for an interview. This callback rate is similar to those achieved in other recent German correspondence tests (Forster & Neugebauer, 2024; Koopmans et al., 2019; Veit & Thijsen, 2021).

3.5 Results

Overall, we find evidence that ethnic majority native applicants are preferred to applicants with a migration background: 52.0% of ethnic majority native applicants receive positive responses compared to only 40.9% of minority applicants (see Table B.3). These results provide support for Hypothesis 1. However, these overall callback rates mask substantial heterogeneity by ethnicity and gender, as shown in Figure 3.2 (left panel). Among male applicants, we find that Danish-heritage men receive about as many callbacks as ethnic majority Germans, but that men of Greek or Turkish origin receive substantially fewer positive responses. While an ethnic majority German man could expect to receive a callback for roughly 1 of every 2 applications (46%), a man with Turkish migration background would need to send more than three applications on average to receive one positive response (a callback rate of 30%), a difference of more than 50 percent. As shown in an exploratory analysis in Figure B.1, Turkish-heritage men also wait longer to receive a response from employers, an average of 8.1 days compared to 5.1 for ethnic majority German men. This pattern suggests that even when Turkish-heritage men are contacted, they may be seen as applicants of last resort.

Turning to female applicants, we find slightly different patterns. First, we find that female candidates are significantly more likely to receive callbacks than male

---

[4]While this number falls short of our pre-registered goal of 3,990 applications, changes to the BA job search website left us unable to collect new job listings as of October 2024.
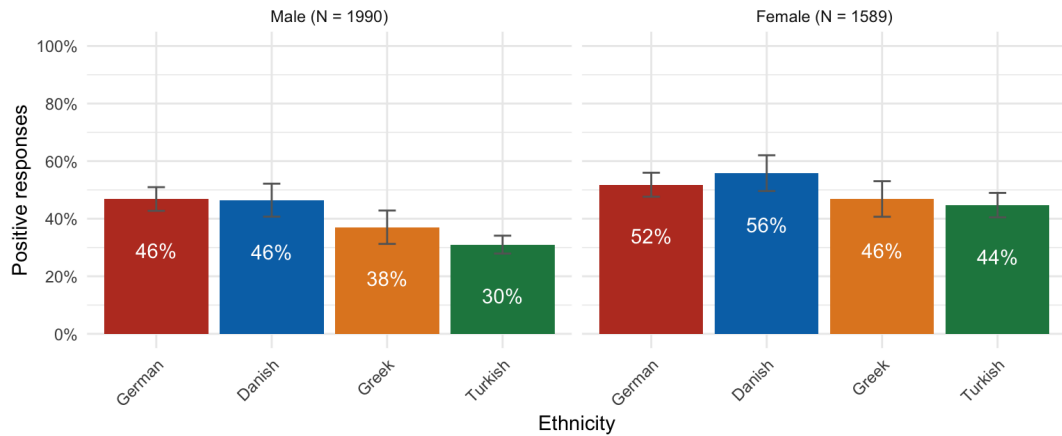
FIGURE 3.2: Callback rates by ethnicity and gender, with 95% confidence intervals. Excluding applicants wearing headscarves (shown in Figure 3.4), $N = 3,579$.
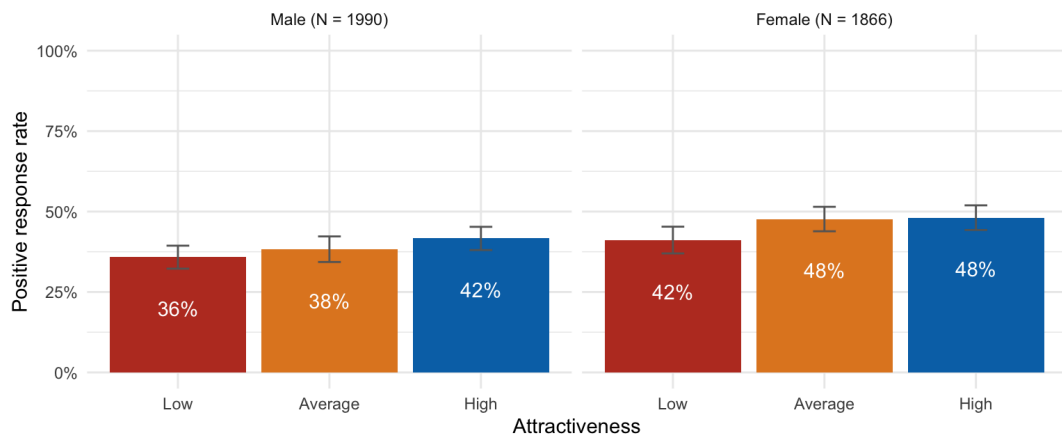
FIGURE 3.3: Callback rates by level of facial attractiveness and gender, with 95% confidence intervals. Full sample, $N = 3,856$.

candidates overall: 47.8% of female candidates vs. 40.6% of male candidates receive positive responses. This does not seem to be driven by a preference for female candidates in part-time jobs, as this preference is stronger in full-time positions (see Figure B.2). We also find gendered patterns with regards to ethnic discrimination. Among women, we do not find significant differences between ethnic majority German women and applicants with Danish or Greek migration background. Only women with Turkish migration background were less likely to receive positive responses, and the difference in callback rates is only 8 percentage points, compared to 16 percentage points between native German and Turkish-origin men. From these results, it would seem that women are generally preferred, and that ethnic discrimination is less severe among female applicants, at least in these groups.

Next, we explore whether facial attractiveness might also contribute to patterns of labor market inequality. As shown in Figure 3.3, we find evidence that highly attractive men and women are significantly more likely to receive callbacks than the

least attractive candidates, with callback rates for both genders 6 percentage points higher in the high-attractiveness condition than in the low-attractiveness condition. Callback rates for applicants with average attractiveness differ by gender. Average-attractiveness women receive roughly as many callbacks as high-attractiveness women, and significantly more than the least attractive women. For men, the callback rate for average-attractiveness applicants does not differ significantly from either the least or the most attractive candidates. This pattern suggests a potential "plainness penalty" for women but a beauty premium for men. Nonetheless, we find support for Hypothesis 2 in that more attractive candidates are generally preferred.

Our next several hypotheses concern the interplay of ethnicity, gender, and physical attractiveness on hiring outcomes. To test these hypotheses, we estimate logistic regression models (given that our outcome measure is binary, i.e., applications receive either a positive or negative response) and present the results as odds ratios for ease of interpretation. Beginning with gender, we find no significant interaction between attractiveness and gender in the model presented in Table B.4, offering no support for Hypothesis 3.

In Table 3.2, we present results on the interaction of ethnicity and attractiveness in two ways. In the first two models, we present a traditional interaction analysis, which estimates a main effect for ethnicity and attractiveness as well as interaction terms for men and women separately. As our hypotheses focused on closing the ethnic gap, we exclude groups that do not face substantial discrimination, namely Danish applicants (results including Danish applicants are shown in Table B.5). Here, we confirm previous results that suggest a significant negative effect for Turkish men, and a positive effect for highly attractive women, although the effect of attractiveness loses significance for men in these models. In Models 3 and 4, we instead estimate the effect of attractiveness only in combination with ethnicity and gender, in line with an intersectional approach which assumes that these characteristics cannot be considered in isolation. These results show a very similar result, with significant negative effects for Turkish men in all three attractiveness categories, as well as positive effects for highly attractive ethnic majority German women. Overall, these findings offer little support for Hypotheses 4 and 5: Attractiveness does not seem to help close the ethnic gap faced by Turkish-heritage applicants, nor does it help Turkish-heritage men. If anything, attractiveness seems to benefit ethnic majority German women the most, which may instead widen ethnic gaps in callback rates.

Our design varied not only identity-related information, but also our applicants' level of work performance, as operationalized by the presence and quality of employer references included with the application. This allows us to test whether appearance-based discrimination might be driven by stereotypes about our candidates' productivity. Theories of statistical discrimination would suggest that when employers have more information about candidates' work performance, they should discriminate less with respect to other features like ethnicity and attractiveness. However, as shown in Table B.6, we find little evidence to support this assertion, and

	(1) Male applicants	(2) Female applicants	(3) Male applicants	(4) Female applicants
Ethnicity (ref. German majority)				
Greek	0.681	0.923		
	(0.183)	(0.261)		
Turkish	0.585 **	0.736		
	(0.112)	(0.183)		
Attractiveness (ref. low)				
Average attractiveness	1.314	0.952		
	(0.308)	(0.210)		
High attractiveness	1.435	1.652 *		
	(0.282)	(0.374)		
Ethnicity x Attractiveness (ref. German x low attractiveness)				
German # Average attractiveness			1.314	0.952
			(0.308)	(0.210)
German # High attractiveness			1.435	1.652 *
			(0.282)	(0.374)
Greek # Low attractiveness			0.681	0.923
			(0.183)	(0.261)
Greek # Average attractiveness	0.835	1.326	0.747	1.166
	(0.333)	(0.518)	(0.203)	(0.332)
Greek # High attractiveness	1.084	0.478	1.059	0.729
	(0.399)	(0.183)	(0.274)	(0.198)
Turkish # Low attractiveness			0.585 **	0.736
			(0.112)	(0.183)
Turkish # Average attractiveness	0.856	1.345	0.658 *	0.942
	(0.254)	(0.430)	(0.129)	(0.209)
Turkish # High attractiveness	0.719	0.718	0.604 **	0.873
	(0.192)	(0.234)	(0.118)	(0.195)
Intercept	0.787	1.029	0.787	1.029
	(0.114)	(0.174)	(0.114)	(0.174)
Number of observations	1610	1277	1610	1277

TABLE 3.2: Effects of ethnicity and attractiveness on callback rates. Logistic regression results presented as odds ratios. All models exclude Danish-heritage applicants. Models 1 and 3 only male applicants; models 2 and 4 only female applicants, excluding veiled Turkish applicants. Standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.

thus no support for Hypothesis 6. While we find that employers prefer highly attractive candidates and candidates with excellent references, none of the interaction terms are statistically significant. Interestingly, we also find that excellent references have a much larger positive effect for female candidates, as shown in Model 3, while among male candidates alone, reference quality has no significant effect. Comparing rates of positive response by ethnicity and reference quality in Figure B.3, we find that this effect is largely driven by women with Turkish migration background. For
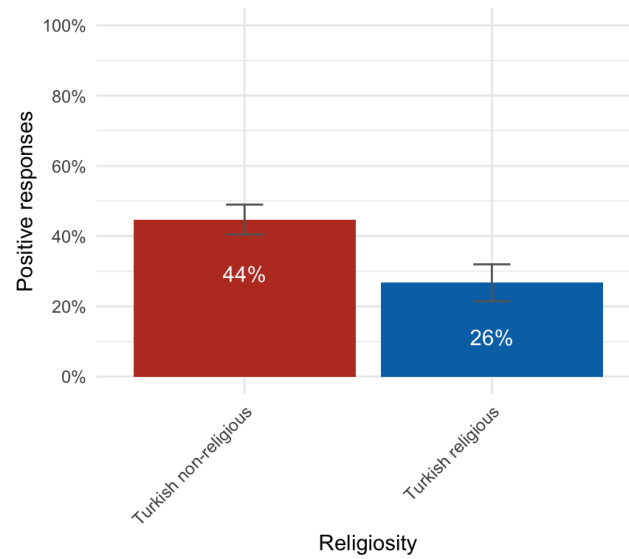
FIGURE 3.4: Callback rates for Turkish-heritage female applicants
with and without headscarves, with 95% confidence intervals. $N =$
807.

men, on the other hand, Turkish-heritage applicants would need excellent references
to achieve roughly the same chance at a callback as an ethnic majority candidate with
no references at all.

Another possible explanation for lookist discrimination on the labor market is
that employers see attractiveness as a productive characteristic. Employers might
expect that their customers will prefer to interact with more attractive employees,
making attractiveness particularly useful in jobs with a high degree of customer con-
tact. We include three such occupations in our test: retail salesperson, receptionist,
and sales associate. In analyses presented in Table B.7, we show that high levels of
facial attractiveness have a positive effect on callback rates, and that callback rates
are generally higher in these three occupations compared to the others. However,
the interaction between high attractiveness and high customer contact is actually
negative (in Model 1, including all applicants), indicating that the value of attrac-
tiveness is lower, not higher, in occupations with extensive customer contact. This
result provides no evidence for our Hypothesis 7, and contradicts previous research
(Rooth, 2009; Tews et al., 2009). As shown in Figure B.4, this effect seems to be driven
by a significant beauty premium for workers in information technology.

Finally, we also test the effect of wearing religious symbols on callback rates.
First, we compare rates of positive response between Turkish female applicants with
and without headscarves, as shown in Figure 3.4. Despite the relatively low levels of
ethnic discrimination shown for women in Figure 3.2, we find substantial discrimi-
nation against Turkish-origin women wearing a headscarf: Only one in four veiled
women receive a positive response, half the rate of ethnic majority native women
and 18 percentage points lower than unveiled women of Turkish heritage. We thus
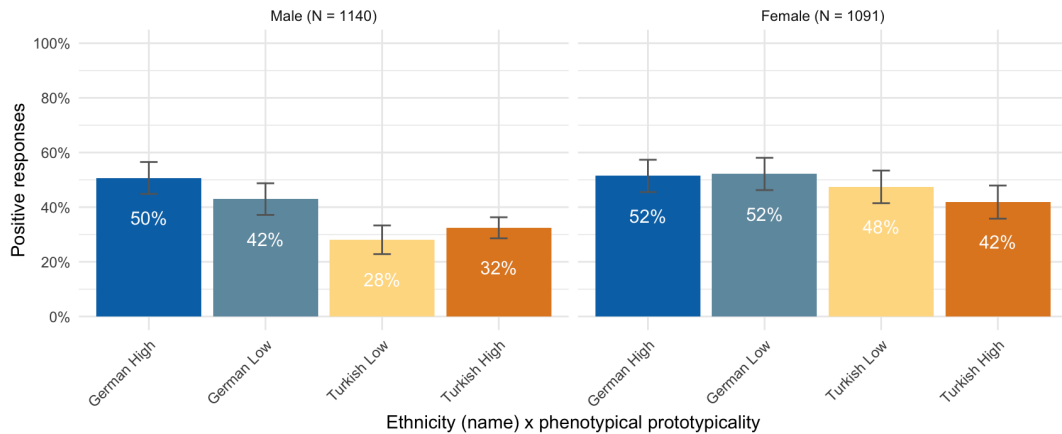
FIGURE 3.5: Callback rates by ethnicity and phenotypical proto-
typicality, with 95% confidence intervals. Includes only applicants
with German or Turkish names, excluding applicants with religious
badges, $N = 2{,}233$.

find strong support for Hypothesis 8. However, turning to the interaction of attrac-
tiveness and religious badges, we find no evidence that wearing a headscarf moder-
ates the effect of attractiveness. In models presented in Table B.8, we find a negative
impact of religious badges for women, but neither the attractiveness coefficients nor
the interactions of attractiveness and religious badges are statistically significant.

3.5.1   Exploratory analysis: names vs. phenotype

In an exploratory analysis, we aim to separate the effects of our two primary eth-
nic signals: name and phenotype. Our selection of photos included some "low-
prototypicality" faces that were seen as plausibly German or Turkish, allowing us to
send applications using the same photo but varying the name (and thus, the per-
ceived ethnicity) of the applicant.[5] In Figure 3.5, we compare callback rates for
the highly prototypical majority German and Turkish-heritage applicants, i.e. those
whose pictures were used only in one ethnic condition, and the less prototypical
applicants. We thus use the same photos in both the low prototypicality condition,
but with ethnically typical names. The observed patterns of positive response mir-
ror those of the test overall (shown in Figure 3.5): There is significant discrimina-
tion against Turkish men, even in the low prototypicality condition. Callback rates
for low-prototypicality profiles are 14 percentage points lower on average when as-
signed a Turkish name, compared to those assigned a German name. This suggests
that ethnic discrimination is not driven by phenotype alone but rather by the percep-
tion of the applicant as Turkish in origin. For women, we observe no significant dif-
ference between low-prototypicality applicants assigned a German name and those
assigned a Turkish name. This again mirrors our earlier results in that ethnic penal-
ties seem to be smaller for women.

---

[5]All photos assigned a Turkish name also listed Turkish as a second native language, which serves
as an additional signal of their ethnicity.

3.6   Discussion

In this study, we conducted a large audit study in the German labor market to explore the combined effects of ethnicity, gender, and physical attractiveness on hiring outcomes. We find robust evidence for an ethnic hierarchy in callback rates: applicants with Danish migration background are treated essentially as native Germans, while men with Greek migration background and all applicants with Turkish migration background face substantial discrimination. We find a difference in callback rates of 16 percentage points between ethnic majority German and Turkish-heritage applicants, a difference that is relatively high compared to other recent work in German-speaking Europe (Di Stasio et al., 2021; Koopmans et al., 2019; Veit & Thijsen, 2021; Zschirnt, 2020). This level of discrimination indicates that even in a relatively tight labor market following the nadir of the Covid-19 pandemic, ethnic discrimination persists (Carlsson et al., 2018). Among women with Turkish heritage, those who wear a veil in their application photo are further penalized. The 18 percentage point difference we find between unveiled and veiled Turkish-heritage women is within the range found in other recent studies (Fernández-Reino et al., 2023; Weichselbaumer, 2020).

Compared to these ethnic and religious gaps, our results for facial attractiveness are modest in size (a difference of only 6 percentage points between the least and most attractive); however, we still find a significant positive effect of attractiveness on the likelihood of being offered an interview. Contrary to our expectations, these effects do not seem to be substantially moderated by ethnicity. However, turning to gender, results from our regression models indicate that women may receive a larger benefit from attractiveness than men. Combining gender and ethnicity, this effect seems to be confined primarily to ethnic majority German women. Thus, physical attractiveness does not seem to close the ethnic gap, but may instead widen it, at least for women.

The ethnic disparities we find are particularly striking given the labor market context in which we conducted our experiment. Comparative research has found that the German labor market exhibits relatively low rates of ethnic discrimination relative to other Western countries (Quillian et al., 2019). This may be due to a relatively unique feature of the German-speaking world: Here, applicants are expected to submit a great deal of information in order to be considered. This information includes scans of educational certificates and references from previous employers, information that should serve as a reliable indicator of productivity. However, even in this highly standardized, high-information setting, we find substantial discrimination against candidates from stigmatized ethnic minority groups, and particularly candidates with Turkish heritage. These findings speak against the theory of statistical discrimination, which posits that employers discriminate on non-productive characteristics only when they lack full information.

Our findings may help inform an emerging debate about "pretty privilege" on

the labor market. While the effects we find for attractiveness are much smaller than the effects we find for ethnicity, we nonetheless find evidence that employers prefer highly attractive applicants. Some jurisdictions, including France; Victoria, Australia; and Washington, DC, have sought to address this problem with public policy by explicitly outlawing discrimination based on a person's appearance (Nault et al., 2020). This debate is perhaps particularly relevant in a setting like Germany, where it is expected to include a photo on one's CV. However, lookism can affect candidates' success at the pre-interview stage even in other contexts due to employers' widespread use of social media to screen potential candidates (Acquisti & Fong, 2020; Baert, 2018). While the use of artificial intelligence in hiring processes may seem like a promising avenue for reducing this inherently human bias, stereotypes about beauty are so deeply ingrained that they may already be encoded in these tools (Matsangidou & Otterbacher, 2019). Even anonymizing applications may only help to reduce bias at the first stage of the hiring process. Given that employers will in almost all cases meet their candidates in person or at least via video conference before making a hiring decision, it may be impossible to keep physical attractiveness out of the decision making process entirely.

Although randomized field experiments remain one of the most reliable means of measuring labor market discrimination, correspondence tests have their limitations. First, we test only at the initial stage of the hiring process, after which we withdraw our applications. It is thus unclear how ethnicity or attractiveness might affect later stages of the application process, where in-person interviews might make both characteristics even more salient to employers. Previous research indicates that discrimination continues well beyond the first interview (Quillian et al., 2020). Second, due to the amount of information needed to apply for jobs in the German labor market, we could test only in a limited number of primarily low- to medium-status occupations. Future research should test in a wider variety of occupations in order to test whether job-specific characteristics (particularly occupational status and occupational sex typing) might impact patterns of discrimination, as suggested by previous research (Marquis et al., 2024; Sarpila et al., 2024). Third, we used a standardized set of photos that kept many potentially relevant appearance-related factors constant in order to identify the effects of attractiveness. Exploring how characteristics like race and weight interact with physical attractiveness may help to uncover additional mechanisms behind ethnic and lookist discrimination (Reece, 2019). Finally, previous studies have shown that contextual factors like the share of migrants in a community or local political attitudes can affect patterns of discrimination (Auspurg et al., 2017; Lacroix et al., 2023). While we test in communities across Germany, we do not have sufficient statistical power to assess how geographic factors might influence labor market discrimination.

Despite these limitations, our results nevertheless have implications for theories of labor market discrimination. The pattern of results presented here is consistent with a conception of attractiveness as a diffuse status characteristic (Webster &

Driskell, 1983). Attractiveness seems to benefit applicants roughly equally regardless of their ethnicity and gender, and thus exerts an additive effect on callback rates, alongside other status characteristics. With respect to the burgeoning literature on intersectionality, we find little evidence that attractiveness has intersectional effects, at least not with regard to ethnicity and gender, in this particular context. Unlike gender and ethnicity, attractiveness does not generally constitute a significant part of a person's identity. It is thus possible that traits like attractiveness are "secondary" traits that do not contribute to stereotypes as strongly as "primary" traits like sex and ethnicity (Heiserman, 2023), a result also suggested by the relatively small beauty premium we find here. However, we are unable to test this question in the present study; more research is needed to determine which traits are relevant to the formation of intersectional stereotypes.

Understanding the mechanisms that drive unequal treatment on the labor market is the first step toward closing persistent ethnic and gender gaps. In this study, we focus on one potential mechanism: physical attractiveness. Our results indicate that despite our theoretical expectations, attractiveness is not a substantial driver of ethnic gaps in the German labor market. We also find that performance information does little to create more equitable labor market outcomes. Even Turkish-heritage applicants with excellent references and educational credentials faced discrimination compared to ethnic majority German counterparts. These results provide some evidence for taste-based rather than statistical discrimination, and suggest that efforts to combat negative stereotypes about minority ethnic groups could be a crucial component of any plan to end labor market discrimination.

Chapter 4

Pretty Trustworthy? Ethnicity, Facial Attractiveness, and Trustworthiness Perceptions

Abstract

Immigrants, and particularly immigrant men, are often stereotyped as untrustworthy in European societies. However, little research has examined how stereotypes of characteristics other than ethnicity might impact natives' perceptions of the trustworthiness of immigrants. Here, I test whether facial attractiveness, a trait associated with a variety of positive stereotypes, might modify ethnic biases in trustworthiness perceptions. I vary facial attractiveness and ethnicity using photo and name stimuli presented in a hypothetical "lost wallet" vignette, in which respondents assess the likelihood of the pictured man returning their lost wallet. Results from an German online panel survey indicate that while attractiveness has a modest positive effect on perceived trustworthiness, the value of attractiveness does not differ between ethnic majority German men and men with a Turkish migration background. Rather, the largest differences in the perceived trustworthiness of Turkish-origin men are found between respondents with inclusionary and exclusionary immigration attitudes, with inclusionary respondents reporting that Turkish-origin vignette persons are more trustworthy than ethnic majority German vignette persons. These results suggest that physical attractiveness does not act as a substantial moderator of ethnic biases in trustworthiness perceptions, but that immigration attitudes are highly relevant.

4.1 Introduction

In contemporary societies, we are often faced with situations where we must place trust in unknown others. In many of these interactions, we must quickly assess a person's trustworthiness based on only a first impression, which is often heavily influenced by the appearance of a person's face (De Neys et al., 2017; Klapper et al.,

---

This chapter has been published in *Scientific Reports* (Hellyer, 2024).

2016; Willis & Todorov, 2006). Whether or not these judgments are accurate (Jaeger et al., 2022), people seem to believe that they can infer the trustworthiness of an interaction partner from their appearance (Eckel & Petrie, 2011) and may rely more on facial cues than information about past performance (Jaeger et al., 2019). Beyond behavior in behavioral tasks, these instantaneous judgments can have serious consequences in a variety of real-world social interactions including partner selection, voting behavior, and criminal justice decisions (Todorov et al., 2015).

A person's face can be used to glean a variety of information, potentially including such traits as ethnicity and gender identity, which might factor into a potential trustor's decision about whether or not to trust. In the language of status characteristics theory, traits like these are considered ascribed status characteristics with levels that are differentially valued (Berger et al., 1972). These particular traits are also "diffuse," meaning that they are associated with general expectations about performance in a variety of tasks, as opposed to a characteristic like language ability that may only affect performance in specific tasks. Thus, stereotypes about the value or prestige of a certain diffuse status characteristic may influence a person's perceived trustworthiness. For example, if members of minority ethnic groups are seen as lower in status than the local ethnic majority, they may also be seen as less likely to behave prosocially in a trust situation. However, empirical evidence for this phenomenon is mixed and may depend on the specific groups and context being studied (Robbins, 2017; Salgado et al., 2021).

From this perspective, one might expect that immigrants, a broadly defined outgroup that commonly faces derogation in many societies, would be seen as untrustworthy. Psychological research on perceptions of outgroup faces (not necessarily immigrants, per se) would support this assertion (Schmid et al., 2022; Sofer et al., 2017). However, experimental evidence on the perceived trustworthiness of immigrants is mixed (Bouckaert & Dhaene, 2004; Cox & Orman, 2015; Gereke & Ruedin, 2023; Kanitsar, 2023). One possible explanation for these findings could be that considerable heterogeneity exists within the broad category of "immigrants," such that certain sub-groups of immigrants are considered particularly untrustworthy while others are not (Lee & Fiske, 2006). For example, recent evidence from Germany finds that men with a migration history are seen as particularly untrustworthy, while women with a migration history are not (Gereke et al., 2020).

Another possible explanation is that considerable heterogeneity exists in natives' attitudes toward immigrants, which might differentially influence their perceptions of immigrants' trustworthiness. While research on immigration attitudes and trustworthiness perceptions is limited, previous research on race in the U.S. has found that implicit race attitudes shape trustworthiness perceptions, with more biased individuals finding Black faces less trustworthy (Hutchings et al., 2024; Stanley et al., 2011). Likewise, we might expect that people with more exclusionary attitudes toward immigrants might find immigrants more threatening and less trustworthy. While recent work finds that preference for natives does not have a significant effect

on trustworthiness perceptions of immigrants (Gereke & Ruedin, 2023), this question has not yet been tested in Germany, or with broader measures of immigration attitudes.

This article seeks to understand how these (gendered) ethnic stereotypes about trustworthiness might also be affected by another highly salient trait: physical attractiveness. Following Dion, Berscheid and Walster's seminal hypothesis of that "what is beautiful is good" (Dion et al., 1972), attractiveness may be a particularly important predictor of perceived moral character (Klebl et al., 2022). Due to the fast and automatic nature of attractiveness judgments (Ritchie et al., 2017), it has been suggested that people use attractiveness as a heuristic to assess the inherently invisible trait of trustworthiness (Gutiérrez-García et al., 2019). This would help to explain results from behavioral games finding that more attractive people are generally thought to be more trustworthy (Pandey & Zayas, 2021; Suo et al., 2024; Wilson & Eckel, 2006; Zhao et al., 2015).

Why might attractiveness affect ethnic stereotypes about trustworthiness? The positive stereotypes associated with physical attractiveness differ substantially from the often negative stereotypes associated with ethnic minority groups. As suggested in the Stereotype Content Model, stereotypes are thought to contain information about two dimensions: warmth and competence (Cuddy et al., 2008). The warmth dimension includes assessments of sociability but also of morality, including perceptions of trustworthiness. Germans rate Turks as lower in warmth than co-ethnics (Froehlich & Schulte, 2019), and gendered stereotypes of criminality, aggression, and misogyny among men from Muslim-majority countries also indicate negative stereotypes about warmth (Wiemers et al., 2024). Attractiveness may provide a signal of warmth or sociability that counteracts some of these negative stereotypes (Eagly et al., 1991), making attractive Turkish-origin men seem more trustworthy than their less attractive counterparts. This would align with previous research that suggests that attractiveness may be a signal of atypicality for members of minority groups (Kunst et al., 2023; Monk et al., 2021), which may be of particular benefit to those in groups facing negative stereotypes, such as Muslim immigrant men. This suggests that attractiveness may benefit men of Turkish descent more than ethnic majority German men.

Based on these findings, I develop several hypotheses about the effects of facial attractiveness and ethnicity on perceptions of trustworthiness. First, I expect to find that highly attractive faces are seen as more trustworthy (H1). In terms of ethnicity, I expect that members of one's ethnic in-group will be seen as more trustworthy. Varying two signals of ethnicity, I expect that either signal of belonging to an ethnic minority group (phenotype in H2 or name in H3) will be associated with lower perceived trustworthiness. Finally, I explore the intersection of facial attractiveness and ethnicity, expecting that attractiveness will have a larger positive effect for ethnic minority men (H4).

To assess whether the size of the "beauty premium" in trustworthiness perceptions varies across ethnic groups, I use a vignette experiment embedded within the German Internet Panel (GIP), a longitudinal study with a large sample that is representative of the German population (Blom et al., 2015). As a measure of the perceived trustworthiness of a vignette person, I use the "lost wallet question," a tool previously used to measure relational trust (Gambetta & Morisi, 2022; Soroka et al., 2007). The term "relational trust" relates to trustworthiness expectations in a two-party framework where one party has an incentive to profit from (or here, withhold assistance from) the other (Hardin, 2002). In the lost wallet question, the respondent is asked to imagine that they have lost their wallet and that it was found by the person described in the vignette. They are then asked to assess how likely they think it is that their wallet will be returned based on the information they are shown about the person. Thus, the measure captures trustworthiness perceptions of a specific person in a specific scenario, improving upon other measures of trust which ask about expectations of a vaguely defined group of "most people" (Landesvatter & Bauer, 2024). Comparing responses between 16 male vignettes that vary in terms of signals of ethnicity (names and phenotypes) and facial attractiveness allows for a comparison of the size of the beauty trust premium across ethnic groups.

4.2   Methods

This article uses data from wave 70 of the German Internet Panel (GIP), a large longitudinal survey of people between the ages of 16 and 75 who live in Germany. Respondents were recruited offline. Initially, respondents from randomly selected neighborhoods were selected via random route sampling and invited to participate in person. In later waves of recruitment, respondents were randomly selected from registry data and invited by mail (Blom et al., 2015). Respondents answer surveys on a variety of topics every two months; wave 70 was fielded in March 2024.

Of the 3,681 respondents who participated in this wave, 1,937 were randomly selected into participating in this experiment. 34 respondents were excluded from this analysis due to missing values for gender, age, and/or the outcome variable. Because respondents' assessment of the vignette person may also be influenced by their own ethnicity, or more specifically whether the vignette person is a member of their ethnic in-group (Schmid et al., 2022; Sofer et al., 2017), I attempt to restrict the sample to ethnic majority native Germans. While no information about ethnicity or migration background is provided in recent waves of the GIP, I use information about respondents' citizenship as the best available proxy. I thus exclude 109 respondents who do not have exclusively German citizenship, leaving me with a final analytical sample of 1,794. As shown in the descriptive statistics provided in Appendix C (Table C.1), the selected sample and the "unselected" sample (those assigned to another experiment) do not differ significantly on most demographic variables used in this analysis. However, once exclusions are applied, respondents

selected for this experiment are slightly more likely to have passed a matriculation examination (*Abitur*), 41% to 38%.

The design of this study was pre-registered on OSF.[1] Each respondent answered one lost wallet vignette containing a headshot-style photo of the person described as well as the following text (translated from German): "[NAME] is 25 years old and grew up near your current place of residence. Imagine that you lose your wallet (containing ID and documents) on the street and [NAME] finds it. What do you think: in this situation, will you get your wallet back?" Responses are given using a 5-point Likert scale ranging from "I will definitely not get it back" to "I will definitely get it back." Given that all vignette persons are said to be raised in the respondents' area, the profiles with migration background should be interpreted as at least 1.5 generation, if not second-generation, immigrants, and not as recent arrivals or tourists. Respondents were randomly assigned to one of sixteen vignettes that vary only in the name and photo shown. However, these two stimuli alone signal several factors, specifically gender, ethnicity, and facial attractiveness.

4.2.1 Stimuli

Face images used as stimuli in this project have been collected from several sources. For Turkish faces, I use a subset of photos from the Bogazici Face Database, a collection of photos of Turkish undergraduate students (Saribay et al., 2018), as well as some photos of Germans with Middle Eastern or North African (MENA) migration history from the DeZIM Picture Database: Faces (Veit & Essien, 2022). For German faces, I use photos of white German men from the DeZIM database and photos of white American men from the Chicago Face Database (Ma et al., 2015). All of the people pictured have explicitly consented to having their photos used in scientific research. All selected faces show a similar expression: neutral or slightly smiling but with no teeth visible. Given that happy faces are seen as more trustworthy (Dong et al., 2014), this selection minimizes the potential for facial expression to significantly affect results. The selected photos have been edited such that all photos have matching clothing and appear against the same background, as shown in Figure 4.1. The edited photos have been rated on their attractiveness, ethnic typicality, perceived social class, and trustworthiness (among other features) by a large sample of German adults from an online access panel (N = 1,125).

Some photos that were rated about equally likely to be Turkish or German will be used in both name conditions, separating the effects of two different cues of ethnicity: the photo and the name. This allows me to test whether visual and name cues have different effects, and whether having an ethnically ambiguous appearance further impacts perceptions of trustworthiness.

I used these ratings, as well as a set of names rated on the same characteristics by another sample of German adults (N = 800), to choose combinations of name

---

[1]https://osf.io/2djpa

FIGURE 4.1: Example vignettes in German low-attractiveness condition (left) and Turkish high-attractiveness condition (right) (German Internet Panel, Universität Mannheim, 2024). Face imagery sources: left, Chicago Face Database (Ma et al., 2015) and right, DeZIM Picture Database: Faces (Veit & Essien, 2022).

and photo that clearly convey gender (male), ethnicity (German or Turkish), and level of facial attractiveness (low or high). To satisfy these conditions, I selected the names Jonas for ethnic majority German men and Mehmet for men of Turkish descent. I also chose photos in order to maximize the difference in attractiveness between groups while minimizing differences within groups, yielding a highly attractive group (mean rating 6.27 out of 11) and a low-attractiveness group (mean rating 4.90).

Altogether, the survey uses 16 different vignettes in a full factorial design. This includes two different photos for each of the 8 unique combinations of attractiveness (high or low), ethnicity (white German or Turkish), and ethnic ambiguity (ambiguous or unambiguous phenotypes). Using two photos helps to ensure that effects are based on the vignette person's characteristics and not driven by possibly idiosyncratic perceptions of a single photo.

4.2.2 Ethical approval

This study design was approved by the University of Mannheim Ethics Commission (EK 04/2023). Informed consent was obtained from all participants of the German Internet Panel, and all participants were reimbursed for their time. The study design complies with German and European law as well as the ethical guidelines of the German Sociological Association.

4.2.3 Statistical analysis

As pre-registered, I first separately estimate the effect of high facial attractiveness on trustworthiness perceptions using OLS models without any additional control variables. Similarly, I estimate the effect of (Turkish) ethnicity on trustworthiness perceptions, separately estimating the effect of a Turkish name among only those vignettes in the ambiguous condition (i.e., those with photos also used in the German name condition) and the effect of a Turkish name and phenotype among vignettes in the unambiguous condition. I then estimate the interaction between attractiveness and (Turkish) ethnicity. Next, I estimate all five of these models with additional control variables that may impact respondents' perceptions of trustworthiness, including respondents' gender (male or female), age (born before or after 1970), educational attainment (*Abitur* or no *Abitur*), employment status (in work or training or not), marital status (living with partner or not), and region (East or West).[2] Statistical analyses were conducted using Stata 17 (StataCorp, 2021).

4.3 Results

Results from regressions without control variables are shown in Figure 4.2, and in tabular form in Appendix C (Table C.2). As hypothesized (H1), I find a positive effect of high attractiveness on trustworthiness perceptions, although this effect is relatively small ($p = 0.037$, Model 1). More surprising are the results for signals of Turkish ethnic background (Models 2 and 3), both of which are associated with significant increases ($p = 0.000$ and $p = 0.002$, respectively) in perceived trustworthiness compared to ethnic German vignettes. This effect appears to be particularly strong among vignettes with unambiguous MENA phenotypical and name signals (Model 2) compared to those with an ambiguous MENA-European phenotype and the same name (Model 3). These findings contradict my expectations in Hypotheses 2 and 3, in which I expected to find a penalty for both signals of ethnic minority group membership.

Finally, in Models 4 and 5, I examine the interplay of facial attractiveness and ethnicity. Despite the overall positive effect of high attractiveness, this effect is only statistically significant for Turkish-origin vignettes ($p = 0.325$ for Jonas and $p = 0.039$ for Mehmet). While this finding points in the expected direction, i.e. that attractiveness may benefit ethnic minority men more than ethnic majority men, analyses shown in Appendix C, Table C.4 show that the value of facial attractiveness does not vary between ethnic groups ($p = 0.399$ without controls). Thus, I find no support for Hypothesis 4.

---

[2]In a deviation from the pre-registration, community size (rural or urban) is not included as a control variable due to missing geographic information for many respondents. Analyses including geographic variables for the sample with valid postal codes are reported in Appendix C, Table C.7.
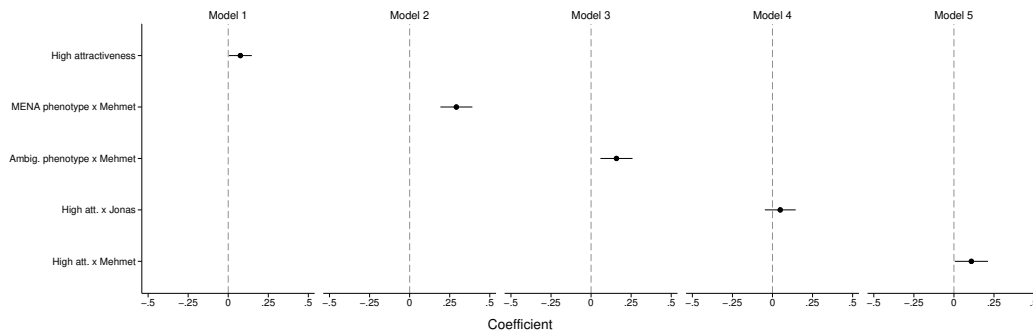
FIGURE 4.2: Coefficients and 95% confidence intervals from multi-variate regression analysis (OLS) without control variables, full sample.

As shown in Appendix C, Figure C.1 and Table C.3, adding control variables does not substantially affect the direction or significance of these results. This demonstrates that these results largely cannot be explained by differences in gender, educational attainment, employment status, marital status, or region. Rather, these variables have effects in the expected directions across treatment categories, such that highly educated, female, older, West German, and partnered people generally assign higher trustworthiness ratings. In a separate exploratory analysis presented in Appendix C, Table C.7, I also add two control variables related to respondents' location of residence (for respondents who reported a valid postal code, $N = 1,370$): the share of migrants in their local area and whether they live in an urban or rural community (defined as a community with fewer than 10,000 residents). These variables, taken from 2011 census data (Budde & Eilers, 2014) (the most recent available with respect to migrant populations), do not exert a significant impact in most models (although respondents in urban areas may be slightly less trusting), and the overall patterns presented above remain unchanged.

Because the outcome variable studied here is ordinal, I also perform an exploratory analysis in which I estimate all models including controls as ordered logistic models. As shown in Appendix C, Table C.5, the same patterns emerge in the results of these ordinal models, suggesting that the choice of estimator is not substantially driving the results. I also perform additional exploratory analyses in Appendix C, Table C.6, that use the mean pre-rated attractiveness value rather than a binary classification of facial attractiveness and focus only on vignettes with unambiguous phenotypes. In these models, the effect of facial attractiveness loses significance ($p = 0.732$ in Model 1). Model 2 suggests that facial attractiveness may have a larger effect for Mehmet, but as in the original analyses, the difference in the effect size of facial attractiveness does not significantly differ between Jonas and Mehmet ($p = 0.569$ in Model 3), even when adding controls ($p = 0.498$ in Model 4). Despite the lack of significant results, the effects are largely in the same direction as previous models, and do not change the conclusion that the effects of facial attractiveness do not differ by ethnicity.

4.3.1 Exploratory analysis: heterogeneous effects

While results regarding attractiveness largely conform to theoretical expectations, the finding of a substantial trust premium for ethnic minority men is wholly unexpected. To explore this in greater detail, I conduct several exploratory analyses to determine whether these overall results might mask heterogeneous results based on other uncontrolled respondent characteristics.

Even though all vignette persons are described as longtime residents of Germany (i.e., they grew up in the area where respondents live), immigration attitudes may be a particularly important characteristic driving attitudes toward vignette persons with a migration background. Respondents with exclusionary immigration attitudes may have stronger and more negative stereotypes of people with migration background than more inclusionary respondents do. Therefore, I subdivide the sample into two groups based on respondents' immigration attitudes. In the same wave of the GIP (after, but not immediately following the trust vignette), respondents answered three questions about immigrants in Germany which have been previously used to measure immigration attitudes in the European Social Survey (Sides & Citrin, 2007; Ziller, 2022). They were asked whether immigration is overall good or bad for the economy, whether immigrants generally undermine or enrich cultural life, and whether it is good or bad that immigrants have the right to come to Germany. Responses were collected using a scale from 0 to 10. I recoded these questions so that positive attitudes were consistently associated with larger values, then calculated the mean of the three responses to create an index of immigration attitudes. I divide the sample into two groups, an "inclusionary" group of those with a score higher than 4 (n = 1,294, 79% of the total sample) and an "exclusionary" group of those with a score of 4 or lower (n = 344, or 21%).

In Figure 4.3 (and Table C.8 in Appendix C), I show results for the ethnicity-related items for each of these two subgroups. Because results are broadly similar between MENA and ambiguous phenotypes, I report results only by name (Mehmet vs. Jonas). As shown in Models 1a and 1b, the positive effect of a Turkish name on trustworthiness perceptions is only found among the inclusionary respondents ($p = 0.000$), while the most exclusionary respondents do not significantly differentiate between German and Turkish vignettes ($p = 0.626$). As shown in Appendix C, Table C.10 (Models 1 and 2), the difference in the effect of a Turkish name between inclusionary and exclusionary respondents is statistically significant ($p = 0.002$ without controls). Additionally, coefficients reported in Table C.8 indicate that exclusionary respondents rate all vignettes as less trustworthy than do inclusionary respondents, suggesting differences in overall trust behavior between these groups. Turning to the interaction of attractiveness and ethnicity, Models 2a and 2b show that inclusionary respondents find more attractive Turkish vignette persons to be more trustworthy relative to their less attractive counterparts ($p = 0.031$), while exclusionary respondents do not differentiate based on attractiveness ($p = 0.806$). This suggests that facial attractiveness cannot close the ethnic "trust gap" for these exclusionary

respondents, but it may serve as an additional signal of trustworthiness for inclu-
sionary respondents rating Turkish profiles.
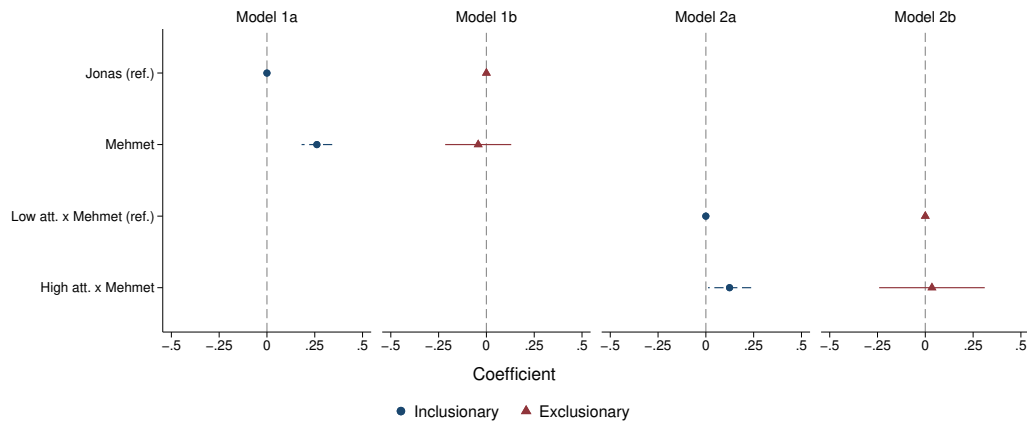
FIGURE 4.3: Coefficients and 95% confidence intervals from multi-
variate regression analysis (OLS) without control variables. Sample
divided into "inclusionary" (index score > 4, n = 1,376) and "exclu-
sionary" respondents (index score <= 4, n = 357).

While the vignettes here do not include any signal of religion or religiosity, re-
spondents may assume that a person with Turkish migration background is Muslim.
As a measure of respondents' attitudes toward Islam, I examine the results of an
experiment fielded in the same wave of the GIP (after but not immediately follow-
ing the trust vignette) that measures respondents' support for building a mosque in
their state's capital city. While the experiment varies textual and visual descriptions
of the planned mosque, I pool the results and group the four response categories
into a single binary indicator of support (n = 734) or opposition (n = 806) to the
construction of a mosque regardless of experimental condition. Results dividing the
sample into these two groups are shown in Appendix C, Table C.9. These results
echo those in Figure 4.3 but with less extreme division between the groups, with
pro-mosque respondents reporting higher perceived trustworthiness for Turkish vi-
gnettes ($p = 0.000$) and anti-mosque respondents reporting no significant difference
between Turkish and German vignettes ($p = 0.227$). As shown in Appendix C, Ta-
ble C.10 (Models 3 and 4), the difference between pro-mosque and anti-mosque re-
spondents' ratings of Turkish-origin vignettes is statistically significant ($p = 0.000$
without controls). Here, attractiveness has no significant effect on trustworthiness
perceptions for Turkish vignettes ($p = 0.136$ for pro-mosque respondents and $p = 0.604$ for anti-mosque respondents), but effects run in the same direction as in Figure
4.3.

I also examine the effect of respondent gender. Considering the effect of gen-
der is important in analyses of physical attractiveness, as (heterosexual) men and
women may respond differently to the attractiveness of male vignette persons. Fur-
thermore, women are more motivated to control prejudice, which may lead them to
report more socially acceptable responses (Harteveld & Ivarsflaten, 2018). However,

there are no substantial gender differences between men and women in these results (Appendix C, Figure C.2 and Tables C.11 and C.12), and while the trends resemble the overall results in Figure 4.2, most coefficients are not statistically distinguishable from zero, with the exception of a large and significant positive effect of a MENA-phenotype Mehmet for both male and female respondents ($p = 0.001$ and $p = 0.000$, respectively), and a significant positive effect ($p = 0.002$) of ambiguous Mehmet for female respondents only.

Finally, I consider the effect of socioeconomic status, which may influence ethnic differences in perceptions of trustworthiness. Previous research has found that the social status of both trustor and trustee may affect trustworthiness perceptions (Salgado et al., 2021). This may be particularly problematic when examining differences related to ethnicity, given that people with Turkish migration background have on average lower educational attainment and income than ethnic majority Germans (Algan et al., 2010). As natives are more likely to discriminate against ethnic minorities who are lower in class (Schaub et al., 2020; Zhang et al., 2019), the results presented so far may conflate ethnic and class-based discrimination. To disentangle these effects, I first examine whether the effects of ethnicity on trustworthiness perceptions vary with respondents' educational attainment, specifically whether or not they received the *Abitur*, as a proxy for overall socioeconomic status. As shown in Appendix C, Table C.13, vignettes with the name Mehmet are seen as more trustworthy by respondents regardless of their educational attainment, even when adding control variables. To estimate the interaction between respondents' socioeconomic status and the perceived status of the vignette person, I also estimate models that include ratings of each vignette person's social class from the pre-test (N = 1,125). In Table C.14 in Appendix C, I show that pre-rated social class has no significant effect on trustworthiness perceptions (in models without controls, $p = 0.642$ for respondents without *Abitur* and $p = 0.261$ for those with *Abitur*), and the name Mehmet is still associated with higher perceived trustworthiness. From these analyses, I conclude that perceptions of class do not seem to underlie respondents' perceptions of Turkish-origin vignette persons.

4.4 Discussion

Results from a vignette experiment confirm previous results about the link between attractiveness and trustworthiness: more attractive profiles were rated as more trustworthy, although the effect size was relatively small. However, despite theoretical expectations, the size of the beauty premium in trustworthiness perceptions does not seem to vary between native ethnic majority German men and men of Turkish descent. While prior research suggested that attractiveness might signal atypicality, reducing negative stereotypes of stigmatized groups like Turkish immigrant men, that does not seem to be the case with respect to trustworthiness perceptions in the

German context. Finally, I find that while respondents with exclusionary immigration attitudes rate members of both ethnic groups as roughly equally trustworthy, those with inclusionary attitudes rate Turkish vignettes as more trustworthy than German vignettes.

This last finding is especially striking, given that previous research has largely found that immigrants, and particularly immigrant men, are perceived to be either less trustworthy (Gereke et al., 2020; Kanitsar, 2023) or about equally as trustworthy as natives (Bouckaert & Dhaene, 2004; Cettolin & Suetens, 2019; Cox & Orman, 2015; Gereke & Ruedin, 2023). Additionally, immigrant men of Middle Eastern descent are often stereotyped as particularly untrustworthy in Europe (Wiemers et al., 2024).

One of two main possibilities seems likely to explain this phenomenon. The first possibility is that the respondents truly found the Mehmet vignettes more trustworthy than the Jonas vignettes. Perhaps the specific vignette information presented here influenced perceptions of the vignette people in ways that were not anticipated. One possibility is that respondents applied other stereotypes to Mehmet vignettes that affected their perceptions, such as the commonly held belief that Middle Eastern men are highly religious (Wiemers et al., 2024). Future work should vary signals of religiosity in order to test this explanation. Another difference between the present research and prior work is the use of the lost wallet question as an outcome measure rather than a behavioral game or field experiment. It is possible that men of Turkish descent are thought to be more trustworthy in this specific scenario (returning lost goods) than in a more abstract or interdependent interaction like a trust game.

The second possibility is that respondents' stated preferences do not match their real world expectations or behavior, i.e., that these results are affected by social desirability bias. Especially in an online survey where respondents have been exposed to other questions about migration, they may sense that the question relates to ethnic differences and respond in line with social norms not to discriminate by ethnicity. An important limitation of this research is that the GIP dataset does not yet include any measure of motivation to control prejudice. However, the present research is in several ways a least likely case for social desirability bias, at least when compared to other survey experimental approaches. First, online surveys are generally thought to be more resistant to such biases than survey modes employing an interviewer, as participants can answer questions anonymously and privately (Kreuter et al., 2008). Second, I employ a between-subjects design where respondents see only a single vignette rather than seeing multiple vignettes that vary in potentially sensitive factors like ethnicity. This design decision was made to make the experimental treatment less obvious and thus reduce the potential for social desirability bias (Walzenbach, 2019). Finally, I signal ethnicity primarily through the visual cue of a photograph along with an ethnically typical name. Recent research finds that respondents generally discriminate more when provided with visual rather than textual (i.e., "he is German") cues of ethnicity like those used here (López Ortega & Radojevic, 2024). Future research should seek to determine the extent to which social desirability bias

impacts survey responses about trustworthiness perceptions of outgroup members. It would also be informative to use other experimental methods that are less susceptible to social desirability, such as field experiments, to assess how appearance-related factors affect behavior in other settings.

Both explanations may be influenced by another factor that has become increasingly important in recent years: political polarization. Immigration has been widely discussed in German political debates since the arrival of hundreds of thousands of primarily Syrian refugees to the country in 2015. This extensive media coverage may have contributed to polarizing attitudes, driving moderates to more extreme views on the subject (Schneider-Strawczynski & Valette, 2025), and thus more divergent trustworthiness perceptions. The hyperpartisan political environment may also affect respondents' perceptions of what constitutes a socially desirable response, such that more exclusionary respondents are more likely to misreport their views (Carmines & Nassar, 2021). This pattern is seemingly consistent with the results presented here, which show a marked preference for Turkish-origin vignette persons for inclusionary respondents and no ethnic differences for exclusionary respondents. If exclusionary respondents actually perceive the Mehmet vignettes to be less trustworthy, this would result in an underestimate of the ethnic trust gap. However, as stated above, my data do not allow me to assess the impact of social desirability bias on my results.

The present research has several other limitations. First, the photos used in the vignettes show only the face, unlike real-world conditions (such as a public place) where one would generally see the whole body. Thus, these results should be understood as the effects of facial attractiveness and not necessarily physical attractiveness. The photos vary only in terms of attractiveness and ethnicity, leaving out other potentially relevant factors like facial threat (Brustkern et al., 2021) or face shape (Leger et al., 2023), or perhaps most notably gender. Future research should assess whether attractiveness alters perceptions more strongly for female faces. Finally, as stereotypes may differ across national contexts, these results may not be generalizable to other countries. More research is needed in order to assess whether these trends hold elsewhere, and to determine what contextual factors might shape differential perceptions of trustworthiness.

Nevertheless, this project adds to the literature on the drivers of trustworthiness perceptions by exploring the interplay of ethnicity and attractiveness, two highly visible traits that are salient to first impressions. Such research on the role of intersectional stereotypes in shaping trustworthiness judgments remains rare, despite increasing recognition that intersectional social identities can significantly modify the content of ethnic stereotypes (Heiserman, 2023; Wiemers et al., 2024). While it does not seem that physical attractiveness plays a substantial role in modifying ethnic biases in trustworthiness perceptions, other physical features or social identities might be more relevant. Additionally, findings about the role of immigration attitudes in shaping perceptions provide additional evidence that the intersection of vignette

and respondent characteristics is important to understanding impression formation
(Schmid et al., 2022; Valmori et al., 2023; Xie et al., 2019). While examining such a
wide variety of factors can be costly and complicated to implement, such research is
needed in order to understand the complex interplay between trustor, trustee, and
context that underlines trust-based interactions in diverse modern societies.

Chapter 5

Pretty Devout? The Effect of Religious Badges on Trustworthiness Perceptions in Turkey

Abstract

People judge the trustworthiness of strangers based on first impressions, often relying on visible cues. While religiosity is generally associated with greater perceived trustworthiness, the role of religious badges—such as specific clothing and grooming styles—remains understudied in Muslim-majority societies, where such signals are widespread and easily recognizable. This study examines how religious badges influence perceived trustworthiness in Turkey, a predominantly Muslim country where expressions of religiosity are neither mandated nor prohibited. Using a large-scale vignette experiment embedded in a nationally representative face-to-face survey, we find that individuals displaying religious badges are perceived as less trustworthy than those without them. Surprisingly, this effect holds even among highly religious respondents. These findings challenge the assumption that religious symbols reliably signal trustworthiness, highlighting the context-specific nature of religious badges in shaping social perceptions. As religious expression remains contested in many societies, understanding how religious markers influence trust is crucial for fostering social cohesion. This study contributes to a more nuanced understanding of religious signaling and social judgments, particularly in Muslim-majority contexts, and offers broader implications for how visible identity markers influence social interactions and group dynamics.

5.1 Introduction

In contemporary societies, we frequently encounter situations that require placing trust in others, often strangers with whom we have no prior relationship. Trust underpins many forms of social and economic exchange: we must trust that an electrician will install the wiring correctly and safely in our new house, or that a taxi driver will take the most direct route to our destination. Even in everyday situations, we rely on trust - whether asking for directions, or leaving belongings unattended in public places. Thus, social trust - the belief that others will uphold the trust placed in them, is a critical component of social exchange and a fundamental building block of our social world (Gambetta & Morisi, 2022; Przepiorka & Berger, 2017; Romano et al., 2017; Schilke et al., 2021).

In many situations, it is difficult to ascertain whether one's interaction partner can be trusted, in particular when we interact with strangers. If we have no information about the person's past behavior in similar situations, we must rely on heuristic shortcuts to decide who to trust (Schilke et al., 2021). Trust decisions then are often influenced by easily accessible information about a person that can be determined at first glance, such as their gender, age, race or looks (Stanley et al., 2011; Wilson & Eckel, 2011).

One such characteristic that is widely believed to influence expectations of trustworthiness is religiosity (Norenzayan & Shariff, 2008; Shariff & Norenzayan, 2007). There are several theoretical reasons to expect that highly religious people may be perceived to be more trustworthy than their less religious peers. First, many religious people adhere to the belief that their actions are being constantly watched by a supernatural entity and, as a result, they may fear divine retribution for dishonesty or reputational rewards for prosociality (Norenzayan, 2013; Purzycki et al., 2016). Second, religious people may have more opportunities to engage in prosocial behavior due to the involvement of religious communities in charitable work (Kelly et al., 2024). These activities not only signal moral commitment, but also provide tangible experiences in prosocial behavior and emphasize a greater sense of connectedness and interdependence (Graham & Haidt, 2010). Finally, religious communities may attract inherently more prosocial individuals who seek this environment precisely because they wish to engage in these community activities (Aksoy & Wiertz, 2024). By voluntarily associating with religious groups, these individuals reinforce the broader societal perception that religiosity correlates with trustworthiness. Together, these factors suggest that religiosity, whether through actual behavioral differences or socially constructed expectations, plays a pivotal role in shaping perceptions of trustworthiness.

A person's religious belief and degree of religiosity are not visible characteristics. However, many individuals wear religious badges or physical markers of religious group membership, which signal affiliation with a specific religious community or

tradition (Sosis, 2005). While intrinsic religious motivations often underlie this practice, they may also be worn for instrumental reasons. For instance, individuals may wear such badges to signal piety to their religious in-group, conveying that they can be trusted to adhere to shared religious customs (Aksoy & Gambetta, 2016; Patel, 2012). Further, wearing a religious badge may act as a "commitment device," reminding the wearer of these religious norms and of the need to resist the temptation to stray from them (Aksoy & Gambetta, 2016; Carvalho, 2013). In some instances, individuals may wear religious badges not out of deep belief but to gain benefits from policies that favor the devout (Aksoy & Gambetta, 2021), or to secure social advantages. Thus, while wearing a religious badge can be costly, as it may be stigmatized by out-group members (Goffman, 1963), it is not always a reliable indicator of a person's true beliefs.

In this paper, we examine the effects of religious badges on perceptions of trustworthiness in Turkey. Even though the association between religiosity and trustworthiness is often assumed to apply broadly, it remains an open question whether religious badges signal trustworthiness, or whether their effect depends on contextual factors or the traits of the observer, including their personal religiosity. Turkey is a strategic research site for our purposes for the following reasons. It is a Muslim-majority country with a long history of public secularism that is unique in its region. Like other Muslim-majority nations, religious badges like head coverings and facial hair are very common in Turkey. However, unlike more conservative countries like Iran or Afghanistan, veiling is not mandatory for women, nor is it banned, as in Tajikistan. While veiling was banned in Turkish government buildings until 2013, Turkish women are now free to choose whether or not to cover their hair in all public settings. While the majority of Turkish women do choose to wear a hijab, about one-third do not cover their hair (Aksoy & Gambetta, 2021). This diversity of religious practice makes the choice to wear a religious badge a more powerful signal than it would be in a context where all women wear a veil. This makes Turkey an ideal case to test the effect of religious badges in a Muslim-majority context, building on previous research in predominantly Christian and Hindu contexts (D. L. Hall et al., 2015; McCullough et al., 2016; Shaver et al., 2018).

Following studies that suggest a positive relationship between religiosity and perceived trustworthiness (Chuah et al., 2016; Chvaja et al., 2023; Power, 2017a; Tan & Vogel, 2008; Thunström et al., 2021), we hypothesized that people displaying outward religious symbols (i.e., religious badges) would be perceived as more trustworthy than those without such visible markers. However, our results from a vignette experiment, fielded as part of a large-scale face-to-face survey of a nationally representative sample of Turkish adults, provide little support for this expectation. Instead, individuals depicted with religious badges, such as head coverings or beards, were generally perceived as *less* trustworthy than the same individuals without these markers. This effect was especially pronounced for more conservative religious badges, such as the chador or distinctively Islamic beard.

Perceptions also varied by respondents' own religiosity but only for female vignettes: compared to non-religious respondents, religious respondents viewed women wearing a moderately religious badge (i.e., the türban) as more trustworthy vis-à-vis those who did not cover their heads or wore a more conservative religious badge (i.e., the chador). Even among religious respondents, however, male vignettes with any form of religious badge, whether moderate or conservative, and female vignettes with the more conservative badge (i.e., the chador) were perceived as less trustworthy than their counterparts without religious markers.

These findings challenge the assumption of a universal link between religious badges and trustworthiness. Rather, they highlight the complex interplay of respondent characteristics (e.g., religiosity), differences in how religious badges function across genders (e.g., veils versus beards), the specific type of badge, and broader contextual factors such as politics. Our results contribute to ongoing public debates about the role of religion in public life, demonstrating that even in a predominantly religious society, public expressions of religiosity can be socially divisive.

5.2   Setting and research design

We investigate judgments of trustworthiness as reactions to signals of religiosity in a pre-registered[1] face-to-face survey experiment with 2,170 Turkish adults. We employed a full factorial design combining textual and visual information. Respondents evaluated six vignettes presented in printed booklets, with responses manually recorded.

To manipulate religiosity, facial profiles of four women and four men were edited to create three versions of each face, corresponding to three levels of religiosity. We added visual markers of religiosity that are common in the Turkish context: head coverings for women and facial hair for men. For both genders, the non-religious condition included no veil or facial hair. For women, the religious version featured a headscarf (*türban*) that covers the hair, while the devout version displayed a black *chador*, which covers all but the face. For men, the religious version included a full beard with a mustache, and the devout version showed a full beard with a shaved mustache, a style typical among many conservative Muslims. Example photographs for each condition are shown in Figure 5.1.

To assess whether these treatments successfully manipulated perceived religiosity, respondents rated each vignette person's perceived religiosity on a scale from 0 to 10. Results shown in Figure D.1 in Appendix D indicate that both head covers and beards significantly influenced perceived religiosity in the expected direction, with head coverings having a more pronounced effect than beards. These results confirm that the manipulation of the three levels of religiosity was successful for both female and male profiles.

---

[1]The pre-registration for this study is available on OSF: https://osf.io/yd3au

The design also controlled for social class and physical attractiveness, two potential confounders. The text indicated one of six common occupations - three low-status (caregiver, receptionist, and sales clerk) and three high-status (doctor, engineer, and lawyer) - to signal social status. We also included two levels of facial attractiveness (low and high, see Materials & Methods below) to isolate the effects of religiosity on trustworthiness judgments.[2]

For each vignette person, the surveyor asked questions about the respondent's first impressions, which respondents answered on scales from 0 to 10. We measure the perceived trustworthiness of the vignette person using a "lost wallet question" (Soroka et al., 2007), which has been widely used in trust research (Bauer & Freitag, 2018; Gambetta & Morisi, 2022). This question asks (translated from Turkish): "Imagine you lose your wallet in a public place. If this person finds your wallet, what is the probability that they will return it to you with everything in it?" This measure improves upon other commonly used measures of generalized trust in that it provides a common frame of reference: it asks respondents about their expectations of a specific person in a specific situation (Landesvatter & Bauer, 2024; Robbins, 2023). The question places respondents in a scenario where they are vulnerable to exploitation, and the finder has no financial incentive to return the wallet beyond an intrinsic desire to help. Respondents must trust that the vignette person will forego any immediate economic gains associated with keeping the wallet (both in terms of money and time) to adhere to the social norm of returning the wallet. This scenario captures the essence of generalized trust while minimizing social desirability bias by focusing on expected behavior rather than directly asking respondents to assess the vignette person's trustworthiness (Gambetta & Morisi, 2022).

5.2.1 Materials and methods

Respondents were randomly assigned to one of 24 survey booklets, each of which included three female and three male vignette persons, one in each religiosity condition. The order of vignettes in each booklet was randomly chosen, and interviewers presented the vignettes in the order in which they were printed. Of these six profiles, three were pre-rated as highly attractive and three as low in attractiveness. The booklets were designed to ensure that no respondent saw the same face in multiple treatments.

Each vignette included a half-A4-size color photograph of the vignette person, and a brief description varying gender and religiosity as well as social class, and physical attractiveness. The vignette text read: "The person whose photo you see is working as a [caregiver]. We will ask you to evaluate this person."

In addition to the lost wallet question described above, we ask for a rating of how "beautiful" (for women) or "handsome" (for men) they find the vignette person, and ratings of the vignette person's religiosity and political leanings. Finally, we asked

---

[2]Previous research indicates that facial attractiveness influences trustworthiness perceptions (Hellyer, 2024; Wilson & Eckel, 2006).

FIGURE 5.1: Religiosity treatments for male and female vignette persons: profiles without religious markers at left, religious at center (with türban or full beard and mustache), and devout at right (with chador or full beard and trimmed mustache). Treatments applied to sample photos from the Bogazici Face Database (Saribay et al., 2018).

a number of demographic questions to ascertain respondents' degree of religiosity, political attitudes, and socioeconomic status.

5.2.2 Face imagery

Face imagery comes from the Bogazici Face Database, a collection of highly standardized headshot-style photos of Turkish undergraduate students (Saribay et al., 2018). This database includes ratings from a sample of 1,207 Turkish adults of various appearance-related factors, including attractiveness, dominance, masculinity, femininity, and trustworthiness. Using these ratings, we selected a set of 16 photos (8 male and 8 female) that provided the desired variation in terms of physical attractiveness. These photos were then edited to standardize their clothing: we removed all jewelry and replaced their clothing with white collared shirts, with different styles for male and female subjects. This editing was intended to remove any undesired signals of social class that might arise from the photo subjects' clothing. We also cropped out most of the subjects' shirts to place more focus on the face. We then had our selection of edited photos rated again by a sample of 550 Turkish adults drawn from an online access panel, to ensure that these changes did not substantially affect how they are perceived with respect to attractiveness.

From this set of 16 photos, we made a final selection of 8 photos, two for each combination of gender (male or female) and physical attractiveness (low or high). These photos were edited again by a professional designer to add the religious treatments. Including all 8 faces in three treatments each (without religious markers,

religious, and devout), we used a final set of 24 photos in the vignette experiment.

5.2.3 Sampling and fieldwork

To ensure broad representation of the Turkish adult population, we used a stratified sampling strategy that randomly selected respondents within 119 neighborhoods in 30 provinces across the country, including at least one province within each of Turkey's NUTS 1 regions. We also stratified by community size to ensure the inclusion of small villages (population less than 2,000), small urban areas, and large metropolitan areas. Additionally, we applied quotas for age and gender, surveying nine men and nine women in each neighborhood, three of each in the age categories 18-33, 34-50, and 51 and over.

Fieldwork was completed from June 28 to 30, 2024. Surveyors from the survey company Konda used a random walk strategy to select respondents, inviting respondents in randomly selected homes to participate. They continued to solicit responses until they filled the quotas in their neighborhood, achieving a response rate of 43%. All respondents were provided with a short summary of the survey and offered their consent to participate. Surveyors conducted 2,263 interviews but they removed 93 interviews for data quality concerns. We thus achieved a final sample of 2,170 responses. To verify data quality, we independently checked the representativeness of the sample on key variables against official statistics such as gender and age distributions and voting patterns. Summary statistics for the full sample are shown in Appendix D, Table D.1.

5.2.4 Statistical analysis

To account for our data structure, which consists of multiple vignette ratings nested within respondents and respondents nested within sampling neighborhoods, we fit multilevel regression models with random intercepts for respondents and neighborhoods. Data analysis was conducted using Stata 17 (StataCorp, 2021). All standard errors reported are robust to clustering at the neighborhood level. Robustness checks presented in Table D.7 were estimated using the REGHDFE package (Correia, 2016).

5.3 Results

We first test whether religious badges are positively associated with perceived trustworthiness, as found in other contexts (D. L. Hall et al., 2015; McCullough et al., 2016; Northover et al., 2024; Shaver et al., 2018). Fitting multilevel regression models with random intercepts for respondents and neighborhoods, we find that this relationship is actually negative for male vignettes: as shown in Figure 5.2 and Table D.2 in Appendix D, bearded men are seen as significantly less trustworthy (compared with the profiles without religious markers as reference, religious: β = -0.68, 95% CI [-0.86; -0.50], devout: β = -1.28, 95% CI [-1.47; -1.10]). For female vignettes,
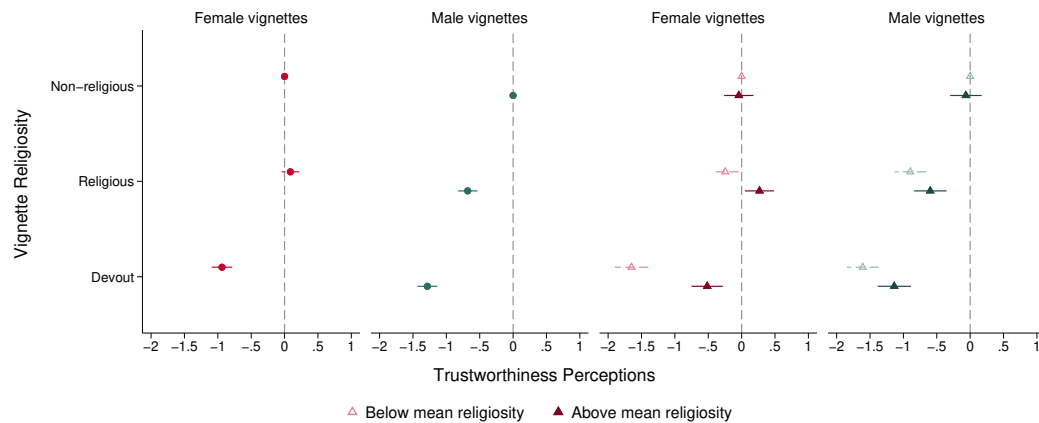
FIGURE 5.2: Left two panels: Perceptions of trustworthiness by vignette gender and religiosity treatment (veil or beard). Right two panels: Perceptions of trustworthiness by vignette gender and religiosity treatment by respondents' self-reported religiosity. Error bars indicate 95% confidence intervals. $N = 12{,}910$ ratings from 2,163 respondents.

the result depends on the type of head covering: women wearing the more common türban are seen as roughly equally trustworthy as those with no head covering ($\beta = 0.09$, 95% CI [-0.06; 0.24]), while those wearing the more conservative chador are seen as significantly less trustworthy ($\beta = -0.94$, 95% CI [-1.16; -0.72]). As shown in Figure D.2, these trends hold regardless of the respondent's gender. Looking at the marginal means in Figure D.8 further highlights these patterns and illustrates the relative effect sizes of religious badges compared to the other experimentally manipulated dimensions of social class and facial attractiveness for male and female profiles.

Next, we examine the role of social class — signaled by the occupation of the vignette person. Including social class as an experimental treatment guards against the concern that any observed link between religiosity and perceived trustworthiness may otherwise be confounded by a potential association between higher religiosity and lower social class (Boon-Falleur et al., 2024; Salgado et al., 2021).[3] Tables D.2 (Models 2 and 6) and D.3 (Model 2) show that when controlling for social class of the vignette person, the main effects of religiosity on perceived trustworthiness remains unchanged.

---

[3]As illustrated in Figure D.3 and Table D.4 in Appendix D, our results also show a positive relationship between the level of social class and perceived trustworthiness ($\beta = 0.71$, 95% CI [0.62, 0.80]). The positive relationship appears consistent across levels of religiosity and gender in the vignettes (see Models 3-4 in Table D.4).

Because the religious treatments may also influence perceptions of facial attractiveness, a known predictor of trustworthiness perceptions[4], we include facial attractiveness, operationalized at two levels (high and low). Our analyses in Tables D.2 (Models 3 and 7) and D.3 (Model 3) show that controlling for facial attractiveness does not alter the main effects of religiosity on trustworthiness perceptions.

To assess whether attractiveness moderates the relationship between religiosity and perceived trustworthiness, we examine their interaction (Figure D.4, Table D.6). We find that highly attractive vignette profiles are rated as more trustworthy in all treatments for male vignettes. However, for female vignettes in the religious and devout conditions, effect sizes do not differ significantly by level of physical attractiveness. These findings suggest that attractiveness is less salient to trustworthiness evaluations of veiled women, potentially indicating that wearing a veil becomes the dominant cue in trustworthiness evaluations.

Next, we explore whether the link between religious badges and perceived trustworthiness varies according to respondent characteristics. We hypothesized that more religious respondents might react more strongly to religious badges, and thus would find more religious vignette persons to be more trustworthy. To test this, we split our sample into respondents above and below the mean value of self-reported religiosity (3.60 on a scale from 1 to 5). Figure 5.2 reports the coefficients for each group. While vignettes wearing religious badges are still generally seen as less trustworthy than those without, we see that more religious respondents rate women in the religious and devout conditions as significantly more trustworthy than do less religious respondents (see Table D.7, Model 2). More religious respondents also rate vignette women wearing the türban as more trustworthy than women without religious markers, in line with our expectations. We find a similar trend for male vignettes in the devout condition, in which more religious respondents find these faces more trustworthy than do less religious respondents. However, we find no significant difference among male vignettes in the religious condition between religious and less religious respondents.

We also fit an additional exploratory model in which we define religiosity by whether or not the respondent reports wearing a veil ($n = 620$, or 56% of women in the sample wear a veil) or beard ($n = 673$, or 66% of men have a beard). Results shown in Appendix D, Table D.8 do not differ substantively from those in Figure 5.2 using self-reported religiosity. We also find similar trends when using interviewer ratings of respondents' religiosity rather than self reports, as shown in Table D.9. Finally, in Table D.10 (Models 4-6), we show that these results are robust to an alternate model specification in which we include sample weights and fixed effects for interviewers.

---

[4]We confirm in Tables D.2 and D.3 in Appendix D that highly attractive vignettes are found to be more trustworthy than less attractive vignettes). While physical attractiveness is sometimes associated with perceptions of higher social class (Schunck, 2016), Table D.5 shows that attractiveness and social class exert additive rather than interacting effects on perceived trustworthiness.

In an exploratory analysis, we examine the role of respondents' political attitudes, which may also influence perceptions of religious people. Turkey has been a staunchly secular country with *laïcité* as one of its foundational principles. The current government, led by President Recep Tayyip Erdoğan and his conservative Justice and Development Party (AK Parti, or AKP), however, has sought to expand the role of Islam in public life (Aksoy & Gambetta, 2021). Supporters of the AKP and its coalition partners may thus have more positive views of the highly religious, or conversely people who follow opposition parties may be more suspicious of the religious. Indeed, our results (shown in Figure D.5 and Table D.11 in Appendix D) provide evidence for this hypothesis: respondents who indicated that they voted for AKP in the last general election ($n = 595$, or 40% of those who voted) assigned higher trustworthiness ratings to women in the religious treatment than did supporters of other parties ($\beta = 0.55$, 95% CI [0.25, 0.85] for AKP voters, $\beta = $-0.20, 95% CI [-0.44, 0.05] for voters of other parties). AKP voters also perceived women in the devout condition to be roughly equally as trustworthy as women without religious markers ($\beta = $-0.09, 95% CI [-0.39, 0.21]). However, for male vignettes, wearing a beard regardless of style is associated with less trustworthiness even among AKP supporters. While trustworthiness ratings of men in the religious condition do not vary significantly by political party, AKP voters perceived men in the devout condition as more trustworthy than did other voters (religious condition: $\beta = $-0.48, 95% CI [-0.75, -0.20] for AKP voters and $\beta = $-0.98, 95% CI [-1.23, -0.74] for other parties. Devout condition: $\beta = $-0.80, 95% CI [-1.12, -0.48] for AKP voters and $\beta = $-1.75, 95% CI [-2.01, -1.48] for other parties).

In additional exploratory analyses, we test whether differences in community size or respondent socio-economic status might drive these results. In Figure D.6 in Appendix D, we first demonstrate that there are no significant differences in trustworthiness perceptions between rural and urban respondents for any combination of vignette gender and religiosity. Next, we test the influence of educational attainment in Figure D.7. We find that respondents with and without a university degree do not differ significantly in their perceptions of female vignettes. However, compared to respondents without a university degree, those with a degree find male vignette persons to be significantly more trustworthy across all religious conditions.

Our results also show that both the perceived religiosity and perceived political orientation of profiles mediate the relationship between religious badges and trustworthiness perceptions. While the effect of perceived religiosity on trustworthiness is positive and statistically significant at the 0.001 level for both female ($\beta = 0.28$, 95% CI [0.21, 0.35]) and male profiles ($\beta = 0.29$, 95% CI [0.23, 0.35]), including perceived religiosity in our main model leads to substantial increases in the magnitude and statistical significance of the religious treatment coefficients, particularly for female profiles (see Table D.12, Model 2 for females and Model 7 for males). Similarly, perceived support for the opposition coalition is statistically significant (at the .05 level for women and .001 level for men), and this changes the coefficients for the

religious treatment. Specifically, the coefficient for females with a turban becomes statistically significant (see Model 3 in Table D.12).

5.4 Discussion

In everyday life, individuals must often decide whom to trust, frequently relying on visual cues and heuristics, especially when faced with limited information. In a large-scale nationwide face-to-face vignette experiment conducted in Turkey, we provide new evidence for the relationship between religiosity and perceptions of trustworthiness. Contrary to our initial expectations, religious badges, such as beards on men and the conservative chador on women do not signal higher trustworthiness. In fact, we observe that individuals with these markers are perceived as less trustworthy than their counterparts without religious badges. Interestingly, women wearing a türban are perceived to be equally as trustworthy as women without religious markers, suggesting that the specific type of religious symbol matters in shaping trust perceptions.

These findings highlight the complexity of the relationship of religiosity, religious badges as signs of religiosity and trustworthiness. Among highly religious respondents and supporters of the pro-Islamic and conservative AK Parti, a trust premium emerges for veiled women but remains limited to those wearing a türban; women in a chador do not benefit from this effect. This suggests that religious signs are not universally interpreted as trust-enhancing, even among those with similar religious affiliations. Perhaps most strikingly, no trust premium exists for men, even among highly religious respondents; in fact, religious respondents and those who support the AK Parti perceive bearded men as less trustworthy than their clean-shaven peers.

These findings challenge prior research that reports a significant positive association between perceived religiosity and expected trustworthiness (Chuah et al., 2016; Chvaja et al., 2023; D. L. Hall et al., 2015; McCullough et al., 2016; Power, 2017a; Tan & Vogel, 2008; Thunström et al., 2021). While some studies suggest that only in-group religious badges are associated with perceived trustworthiness (Northover et al., 2024; Shaver et al., 2018), our findings reveal a trust premium only for women wearing a headscarf among religious respondents. We are the first to find evidence of a null to negative association between religious badges and perceived trustworthiness even within a religious in-group. It challenges the assumption that religiosity, by default, enhances trustworthiness and instead points to the importance of context, including the specific cultural and political environment in which these judgments are made.

One plausible explanation for these findings could be Turkey's unique sociopolitical history, where religious badges carry layered meanings. The rise of the openly pro-Islamic AK Parti, which overturned previous bans on veiling, has likely polarized views of women wearing the türban, associating it with both religious

piety and political identity (Atac & Adler, 2024; Çokgezen, 2022). What was once primarily a sign of religious devotion has increasingly taken on political connotations, a shift supported by our findings that AKP supporters have the most positive views of veiled women. This suggests that in Turkey, the trust premium associated with religious badges is not universally applicable but rather localized within an in-group defined by both religious and political beliefs, rather than religiosity alone. Moreover, this in-group premium only applies to women wearing the türban and not to all religious badges, suggesting that women's religious badges are often more politically charged and tied to the performance of religiosity in public life. Gender norms seem to be important here, with women's veiling practices being more heavily politicized and scrutinized compared to men's religious beards in the public sphere.

Social identity theory (Tajfel & Turner, 1979) would suggest that people place more trust in members of the in-group than in members of the out-group (Platow et al., 2012). It has been postulated that religious cues shape expectations about reciprocity based on whether the person is perceived to belong to a shared in-group (Romano et al., 2017; Tanis & Postmes, 2005). However, our results, which consider both respondents' religiosity and the vignette profile's religiosity, challenge a simplistic notion of identity. Even among religious in-group members, perceptions of trustworthiness are negatively affected by religious badges. It seems to be the combination of religion *and* politics that defines in-group membership in the Turkish context. Indeed, research shows that in Turkey, prejudice based solely on religiosity is relatively rare (Aytaç & Çarkoğlu, 2019).

Another related explanation for our results could be found in signaling theory (Gambetta, 2011; Spence, 2002; Zahavi, 1975). In the Turkish context, religious badges such as headscarves and beards may not be sufficiently "costly" to function as reliable signals of trustworthiness. Indeed, given the long-standing dominance of the religious AK Parti for over two decades, the absence of such badges may, in fact, carry greater signaling value. Our results could thus be interpreted as evidence of a secular trust premium rather than a religious trust penalty.

A final theory which may be relevant for our study is moral licensing (J. C. Jackson & Gray, 2019; Merritt et al., 2010; Simpson et al., 2013). This theory suggests that individuals who signal adherence to higher moral or divine religious standards may feel licensed to engage in morally questionable behaviors in more mundane affairs. Religious badges may thus fail to universally enhance trustworthiness because observers might interpret such displays of religiosity as forms of moral compensation for questionable practices in mundane situations such as when finding a lost wallet. This could be especially true in a polarized socio-political environment, where religious symbols are intertwined with political identity and potential strategic signaling. While moral licensing offers a potential explanation for some of the observed patterns - such as the differential trust premium for women wearing

a türban versus no headscarf, it does not fully account for all of our findings. Additional mechanisms, including context-specific interpretations of religious badges, likely contribute to the findings.

We also uncover that the relationship between religious badges and perceived trustworthiness is highly gendered. Specifically, we observe that veiling as a religious marker for women functions as a more distinctive signal than beards for men (see Figure D.1 in Appendix D). While the türban was associated with higher perceived trustworthiness among more religious respondents or AKP supporters, beards showed no such effect in any group, even though they too signal religiosity. Our findings may indicate that while bearded individuals are perceived to be more religious, beards may also signal other attributes, such as masculinity and potential threat (B. J. W. Dixson et al., 2021), which could negatively impact perceptions of trustworthiness beyond religiosity. Moreover, we observe that bearded profiles are rated as less attractive, and perceived attractiveness may serve as a mediator in the relationship between religious markers and trustworthiness. In fact, when we control for perceived attractiveness, the negative effects associated with religious markers become non-significant, suggesting that attractiveness plays a key role in shaping trustworthiness perceptions linked to these religious badges (see Model 9 and 10 in Table D.12).

Like all studies, ours has some limitations. While we show that religious badges are associated with higher perceived religiosity and that our experimental design controls for social status and facial attractiveness, we acknowledge that social status encompasses more than just occupation and may still play a role in how these religious markers are interpreted. Additionally, while the badges signal religiosity and political orientation, they likely convey more complex cues that we have not fully captured. Our study focuses on veils and beards as markers of religiosity, but future research should consider whether different signals of religiosity might yield different results. It would also be valuable to examine the context-dependent stereotypes associated with various religious badges. To gauge the generalizability of our findings, future studies should replicate our experiment in a variety of contexts and trust situations. Turkey is not necessarily representative of other Muslim-majority countries, especially countries where veiling is mandatory. Nor is it likely to be representative of countries where Muslims are a religious minority, such as many European countries, in which displaying Islamic religious badges may be controversial or more costly (Henrich, 2009). Testing badges from multiple religious traditions in various contexts would also help answer the question of whether religious badges signal trustworthiness only to religious in-group members, as suggested in previous work (Northover et al., 2024; Shaver et al., 2018) or whether the badge-trustworthiness link strengthens with the costliness of displaying the badge, as signaling theory would suggest (Gambetta, 2011; Spence, 2002; Zahavi, 1975).

Our results suggest that the association between religiosity, religious badges and perceived trustworthiness is context-dependent and shaped by societal and cultural

factors. Similar to how the effects of religiosity on prosocial behavior diminish and ultimately disappear in contexts with high levels of social enforcement of religiosity, the association between religious badges and trustworthiness perceptions may shift based on the political and cultural environment (Stavrova & Siegers, 2014). Future work should explore the mechanisms driving these perceptions, examining how factors such as political climate, social norms, and individual experiences influence the perceptions of religious symbols. While past work suggests that signs of religiosity should be associated with positive social evaluations, our study shows that under certain socio-political conditions, these signs may completely switch in their signaling content and in fact be interpreted negatively. This highlights the importance of considering how both the receiver's characteristics and contextual factors influence the interpretation of visible cues and their effects on social perception.

Chapter 6

Discussion & Conclusion

In this dissertation, I conducted four studies that investigate how appearance shapes both interpersonal perceptions and labor market outcomes. These studies tested the combined effects of a variety of traits conveyed primarily through visual cues, including ethnicity, religiosity, gender, and physical attractiveness. In particular, I focused on the question of whether physical attractiveness, a characteristic that is associated with a number of positive stereotypes, could counteract the negative stereotypes associated with minority ethnic and religious groups, thus helping to offset ethno-religious disparities. Although this idea has some theoretical support from status characteristics theory and Bourdieu's theory of capital, I found little evidence for this hypothesis. Attractiveness did not consistently have a larger effect for members of stigmatized groups. In fact, attractiveness more often benefited members of high-status groups, at least in the contexts studied here.

In the first two studies of my dissertation, I focused on labor market outcomes. In Chapter 2, my co-authors and I analyzed data from pairfam, a large, longitudinal study of German families that also includes measures of respondent attractiveness, to assess whether ethnicity affected the size of the beauty premium in the German labor market. Estimating hundreds of models in a multiverse analysis, we found evidence that physical attractiveness provides a more consistent benefit for ethnic majority native German men and women than it does for members of minority groups. Results for minority ethnic groups were highly inconsistent, offering little evidence of a significant beauty premium for members of these groups. While some models showed a significant beauty premium for women with Turkish migration background as well, this effect was not statistically significant in about half of the models tested. These results offered little support for the idea that attractiveness might close ethnic gaps in earnings. Instead, we find that attractiveness may even exacerbate these persistent wage inequalities.

Next, in Chapter 3, I turned my attention to the hiring context, conducting a labor market correspondence test in which I sent nearly 4,000 job applications to real job postings across Germany. Using application photos added to résumés, as is customary in Germany, my co-authors and I varied our fictional applicants' gender, ethnicity, religiosity, and level of physical attractiveness, in order to test how these effects combine in terms of positive callbacks. Here, we find strong evidence

for a gendered ethnic hierarchy. Men of Greek and Turkish heritage face substantial penalties in callback rates compared to ethnic majority Germans or Danish-heritage applicants. For women, only Turkish-heritage applicants face a significant ethnic penalty, which is also smaller in size than the penalty for men. One exception is that women who wear a veil, signaling adherence to Islam, face substantial discrimination, receiving half as many callbacks as ethnic majority German women. We also find evidence that employers prefer more attractive candidates, although the size of this effect is generally smaller than for ethnicity. We turn to regression models to test the interaction between attractiveness and ethnicity, finding a significant beauty premium only for ethnic majority German women. As in Chapter 2, this evidence points to a larger beauty premium for the ethnic majority, leading to widening rather than shrinking ethnic gaps at higher levels of physical attractiveness.

In the latter half of my dissertation, I focused on trustworthiness perceptions, which are important drivers of interpersonal behavior and possibly also moderators of the labor market disparities I find in Chapters 2 and 3. In Chapter 4, I begin by exploring how physical attractiveness and ethnicity might combine in shaping trustworthiness perceptions of young men in the German context. I test this question with a vignette experiment embedded in a large, representative online panel, in which respondents see a single face and estimate how likely they believe it is that the pictured person would return their wallet to them if it were lost. I hypothesized that I would find a substantial ethnic penalty in trustworthiness perceptions, but that highly attractive men with Turkish migration background would be treated more like the ethnic majority. Instead, I find a significant trust *premium* for Turkish-heritage men, rather than the hypothesized penalty. While I also find a trust premium for highly attractive men, the size of this premium does not vary significantly with ethnicity. Rather than attractiveness, the pattern I find seems to be driven by respondent immigration attitudes: more inclusionary respondents report more trust in Turkish-heritage men, while more exclusionary respondents report about equal perceptions of trustworthiness across ethnic groups. Unlike previous chapters, here I find no significant difference in the effects of attractiveness across ethnic groups.

Finally, in Chapter 5, I investigated perceptions of trustworthiness in the Turkish context, focusing on groups defined by religiosity rather than ethnicity. We tested the effects of religiosity and attractiveness on trustworthiness using lost wallet vignettes, as in the previous study. We included six such vignettes in a face-to-face survey of more than 2,000 adults across Turkey, conducted by a team of trained surveyors in respondents' homes. We conveyed information about each vignette person's religiosity using common religious badges worn in Turkey: two types of headscarves for women and two types of beards for men. While we expected to find that religiosity had a strong positive effect on trustworthiness perceptions, we found the opposite: men with religious beards and women wearing the more conservative *chador* head covering were seen as less trustworthy than those wearing no badge. This is a fully unexpected result, especially considering that these patterns largely

hold even for devout Muslim respondents. In terms of the interaction between religiosity and attractiveness, we find that attractiveness benefits all applicants except for veiled women, suggesting either that the veil suppresses the size of the beauty premium or that religiosity becomes a more dominant cue in trustworthiness perceptions for these religious women.

In summary, I find that physical attractiveness, ethnicity, and religiosity contribute to both trustworthiness perceptions and labor market outcomes. Ethnicity and religiosity seem to have particularly large effects on both outcomes, while attractiveness plays a smaller but still significant role in most cases. This overall trend provides some evidence for the idea that some social categories act as "master statuses" that have particularly strong effects on interpersonal perceptions (Goffman, 1963). The traits usually suggested as master statuses tend to be core aspects of identity, such as age, gender, and place of origin (Neuberg & Sng, 2013). Factors like gender, ethnicity, and religion may thus be more central to the formation of stereotypes than is attractiveness. This organization of information may relate to the order in which we perceive different traits: research suggests that we process information about gender before information about physical attractiveness (Carbon et al., 2018).

However, the results of the studies presented here did not yield a clear pattern as to how these variables interact. Generally, attractiveness seems to have a larger benefit for ethnic majority natives in the labor market context, while it seems to provide roughly equal benefits with respect to trustworthiness. One notable exception is that veiled women in Turkey receive a much smaller beauty premium with respect to trustworthiness perceptions, suggesting that attractiveness is a less relevant cue than religiosity with respect to trust, and/or that covering the head strongly reduces the effects of attractiveness. The conflicting patterns found here may relate to the intersection of context-specific stereotypes, which should be investigated in greater detail in future studies.

Returning to the theoretical frameworks discussed in Chapter 1, this pattern of results would tend to support an interpretation based on social identity theory, at least in terms of labor market outcomes. Attractiveness seems to offer a larger and more consistent benefit to ethnic majority group members in terms of wages and interview callbacks. This may suggest that from the perspective of decision makers who are likely to be members of the ethnic majority group, highly attractive ingroup members are preferred. However, research on the "black sheep effect" would also suggest particularly strong penalties for unattractive ingroup members in order to maximize group distinctiveness (Marques & Paez, 1994). This does not match the callback patterns we find in Chapter 3. Rather, we find that even the least attractive ethnic majority applicants are still preferred to even highly attractive Turkish-heritage applicants.

This pattern of results is also consistent with research on the outgroup homogeneity effect, which suggests that people are less able to differentiate between the

traits of outgroup members (Boldry et al., 2007; Ostrom & Sedikides, 1992). If decision makers assume that all outgroup members are the same, both in terms of appearance and other characteristics, then they may not perceive differences in attractiveness as readily as they do for ingroup faces. However, evidence for this effect is mixed (Zebrowitz et al., 1993), and my own results from the two studies on trustworthiness suggest that attractiveness also has an effect on perceptions of members of ethnic and religious outgroups, contradicting the idea that differences in attractiveness are imperceptible to outgroup members.

Findings from the two studies on trustworthiness perceptions also highlight that respondent characteristics have a large influence on how certain visual cues are perceived. In the German study in Chapter 4, immigration attitudes had a large effect on the perceived trustworthiness of Turkish-heritage men. In the Turkish study in Chapter 5, respondents' religiosity played a large role in their perceptions of the trustworthiness of vignette persons wearing religious badges. This too suggests a pattern consistent with social identity theory in that the boundary between ingroup and outgroup seems to be highly consequential. Such contextual factors are also likely to matter for labor market outcomes as well, but I was unable to study this in depth due to the very limited information available about decision makers in my labor market studies. Future research should endeavor to study how contextual factors influence appearance-based bias on the labor market.

The work presented here also raises other important questions for future research on the influence of visual cues. One key limitation of the present research is that I was only able to study two types of outcome variables, labor market outcomes and trustworthiness perceptions. While both of these outcomes are important in driving patterns of ethnic, gender, and religious inequality, they are by no means the only venues where unequal treatment is prevalent. In particular, research has already established beauty premia in education (Bauldry et al., 2016; Umberson & Hughes, 1987), as well as combined effects of attractiveness and race (Parks & Kennedy, 2007). As an evolutionary sign of sexual fitness, physical attractiveness is also likely to have a strong effect on outcomes in the partner market (Rhodes et al., 2005). It could be interesting to explore whether these positive effects can counteract ethnic inequalities in educational attainment (Kristen & Granato, 2007), or preferences for ethnic and religious homophily in romantic relationships (Carol, 2018; Heyne et al., 2025).

Future work should also investigate the effects of a wider variety of visual cues. As noted in Chapter 1, appearance is an extremely broad cue that can signal many different traits. Here, I studied the effects of highly salient cues such as gender, ethnicity, religiosity, and physical attractiveness. In order to maintain sufficient power to test these variables, other important traits like age and body weight were held constant. Age in particular has been noted as a core element of interpersonal perception (Neuberg & Sng, 2013; Penner & Saperstein, 2013), and weight has also linked to significant inequalities in the labor market (Goulão et al., 2024; Rooth, 2009). The present work can only be generalized to the relatively young (roughly college-age)

faces studied here, which are also generally average in terms of weight. Future research may wish to vary these elements in order to test how they might play into perceptions of attractiveness and also resulting inequalities.

Another highly salient visual cue left unexplored here is race. This omission may help to explain the differences between the results I find here and previous (mostly American) research on this topic (such as Kunst et al., 2023; Monk et al., 2021). It has long been assumed that race is less relevant in the European context than it is in the U.S., where racial divisions have been deeply entrenched for centuries. However, despite the paucity of research on European racism, racial discrimination seems to be prevalent and distressing for people of color in Germany (Scholaske, 2024; Williams, 2024). Given that racial differences may be even more salient markers of foreignness than the sometimes ambiguous cues of ethnicity I study here (Gereke et al., 2022), studying the interaction between race and attractiveness may yield a pattern of results more like that found in the United States.

Due to my focus on visual cues, I also neglect other sensory information that may influence first impressions. For example, accented speech may provide an additional signal of foreignness that exacerbates ethnic biases on the labor market (Schmaus & Kristen, 2022). Previous work has already found that accent-based bias is affected by whether a person belongs to a visible or invisible minority group (Hansen et al., 2018). Such work highlights the potential for studying the interplay between visual and auditory cues with respect to first impressions, especially on the labor market where native-sounding speech may be particularly favored.

Despite these limitations, I find significant evidence that appearance matters for both trustworthiness perceptions and labor market outcomes. We may know deep down that a person's appearance is not a reliable measure of their character, yet we still rely on appearance to guide our perceptions of and behavior toward other people. In the aggregate, these instantaneous judgments can lead to discriminatory patterns of behavior that can do lasting harm to members of visible minority groups. The patterns presented here show that the interactions between visual cues like ethnicity and attractiveness are remarkably complex. Studying these interactions in greater detail will require extremely large samples and more varied stimuli that can test a wider variety of characteristics. However, the results presented here show the potential of such research for developing a richer understanding of interpersonal perception. This is necessary work if we hope to disrupt the persistent patterns of stereotyping and discrimination found in our increasingly diverse societies.

Appendix A

Appendix to Chapter 2

A.1   Descriptive results

Figure A.1 shows the distribution of interviewers' ratings of physical attractiveness separately for women and men, and also by ethnicity. While women of almost all ethnicities are most likely to be classified as highly attractive, most men are rated a 6 out of 7. Among women, multiethnic women are most frequently rated as being very attractive, followed by the mixed category of other non-German, one third of whom are rated as being very attractive. Among men, natives followed by Ethnic German immigrant men are most often rated as being very attractive, while other non-Germans and those with a Turkish background are least often rated as being very attractive.
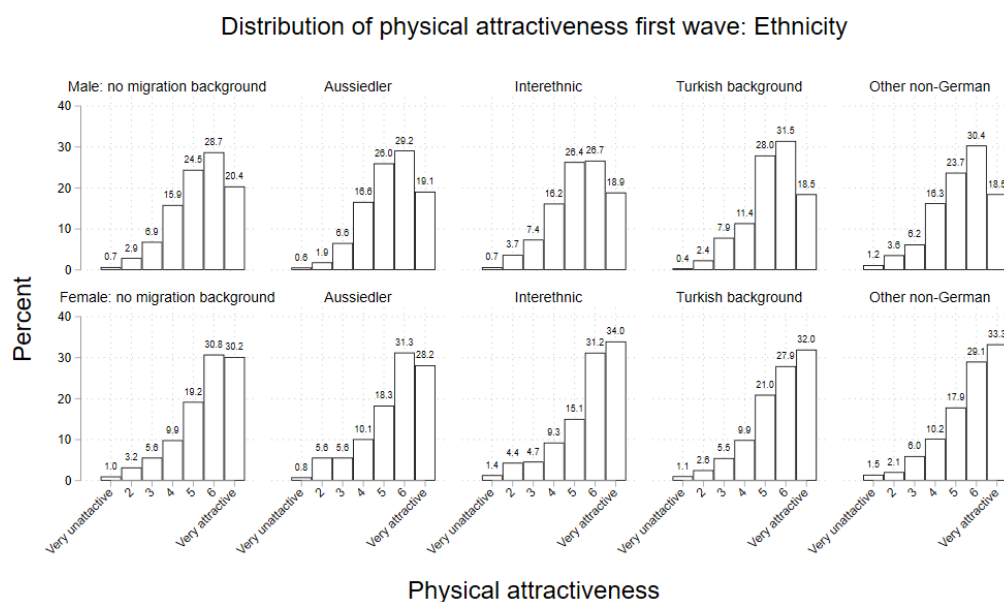
FIGURE A.1: Distribution of pairfam interviewer ratings of physical attractiveness by ethnicity.

TABLE A.1: Women: sample characteristics for preferred specification
(N = 7,420).

	mean	sd	count
Native German			
Log hourly earnings, trimmed 5% / 95% quant.	2.44	0.51	5929
Attractiveness rating, wave 1, bin.	0.30		5929
Attractiveness rating, wave 1, cont.	5.58	1.40	5929
Age	33.13	7.97	5929
Sample: Cohort 1	0.24		5929
Sample: Cohort 2	0.33		5929
Sample: Cohort 3	0.43		5929
Currently enrolled	0.10		5929
No degree/lower secondary	0.04		5929
Upper secondary, vocational	0.40		5929
Upper sec. general, post sec.	0.15		5929
Tertiary	0.31		5929
Parent. No degree/lower secondary	0.04		5929
Parent. Upper secondary, vocational	0.62		5929
Parent. Upper sec. general, post sec.	0.05		5929
Parent. Tertiary	0.29		5929
B5 - extraversion, wave 1, factor score	-0.21	0.84	5929
B5 - agreeableness, wave 1, factor score	-0.11	0.81	5929
B5 - conscientiousness, wave 1, factor score	0.23	0.72	5929
B5 - neuroticism, wave 1, factor score	0.07	0.83	5929
B5 - openness, wave 1, factor score	0.04	0.76	5929
Self-rated health	3.64	0.94	5929
BMI, wave 1	23.76	5.04	5929
Interview duration	57.81	17.44	5929
Height (cm), wave 1	168.36	6.31	5929
Language ability, wave 1	3.98	0.21	5929
German citizenship, wave 1	1.00		5929
Religiosity, wave 1	1.63	1.41	5929
Ethnic German ("Aussiedler")			
Log hourly earnings, trimmed 5% / 95% quant.	2.40	0.46	368
Attractiveness rating, wave 1, bin.	0.27		368
Attractiveness rating, wave 1, cont.	5.51	1.37	368
Age	31.98	7.59	368
Sample: Cohort 1	0.24		368
Sample: Cohort 2	0.44		368
Sample: Cohort 3	0.32		368
Currently enrolled	0.12		368
No degree/lower secondary	0.08		368

Upper secondary, vocational	0.35		368
Upper sec. general, post sec.	0.13		368
Tertiary	0.32		368
Parent. No degree/lower secondary	0.06		368
Parent. Upper secondary, vocational	0.40		368
Parent. Upper sec. general, post sec.	0.11		368
Parent. Tertiary	0.43		368
B5 - extraversion, wave 1, factor score	0.08	0.86	368
B5 - agreeableness, wave 1, factor score	-0.09	0.78	368
B5 - conscientiousness, wave 1, factor score	0.15	0.85	368
B5 - neuroticism, wave 1, factor score	0.22	0.73	368
B5 - openness, wave 1, factor score	0.09	0.79	368
Self-rated health	3.62	0.94	368
BMI, wave 1	23.53	5.23	368
Interview duration	59.63	18.43	368
Height (cm), wave 1	165.51	6.15	368
Language ability, wave 1	3.90	0.29	368
German citizenship, wave 1	1.00		368
Religiosity, wave 1	2.07	1.23	368
Multiethnic ("Half-German")			
Log hourly earnings, trimmed 5% / 95% quant.	2.47	0.53	450
Attractiveness rating, wave 1, bin.	0.31		450
Attractiveness rating, wave 1, cont.	5.76	1.33	450
Age	34.45	7.77	450
Sample: Cohort 1	0.17		450
Sample: Cohort 2	0.31		450
Sample: Cohort 3	0.52		450
Currently enrolled	0.09		450
No degree/lower secondary	0.04		450
Upper secondary, vocational	0.30		450
Upper sec. general, post sec.	0.18		450
Tertiary	0.40		450
Parent. No degree/lower secondary	0.09		450
Parent. Upper secondary, vocational	0.47		450
Parent. Upper sec. general, post sec.	0.05		450
Parent. Tertiary	0.39		450
B5 - extraversion, wave 1, factor score	-0.31	0.79	450
B5 - agreeableness, wave 1, factor score	-0.10	0.83	450
B5 - conscientiousness, wave 1, factor score	0.11	0.80	450
B5 - neuroticism, wave 1, factor score	0.05	0.87	450
B5 - openness, wave 1, factor score	0.21	0.73	450

Self-rated health	3.62	0.96	450
BMI, wave 1	23.47	3.92	450
Interview duration	60.45	20.22	450
Height (cm), wave 1	167.09	5.71	450
Language ability, wave 1	3.99	0.11	450
German citizenship, wave 1	0.97		450
Religiosity, wave 1	1.67	1.38	450
Turkish background			
Log hourly earnings, trimmed 5% / 95% quant.	2.26	0.49	152
Attractiveness rating, wave 1, bin.	0.33		152
Attractiveness rating, wave 1, cont.	5.41	1.53	152
Age	32.83	7.83	152
Sample: Cohort 1	0.24		152
Sample: Cohort 2	0.32		152
Sample: Cohort 3	0.43		152
Currently enrolled	0.11		152
No degree/lower secondary	0.31		152
Upper secondary, vocational	0.32		152
Upper sec. general, post sec.	0.15		152
Tertiary	0.11		152
Parent. No degree/lower secondary	0.74		152
Parent. Upper secondary, vocational	0.20		152
Parent. Upper sec. general, post sec.	0.03		152
Parent. Tertiary	0.02		152
B5 - extraversion, wave 1, factor score	-0.13	0.85	152
B5 - agreeableness, wave 1, factor score	-0.30	0.81	152
B5 - conscientiousness, wave 1, factor score	0.26	0.74	152
B5 - neuroticism, wave 1, factor score	-0.09	0.78	152
B5 - openness, wave 1, factor score	-0.08	0.68	152
Self-rated health	3.48	0.99	152
BMI, wave 1	25.44	4.17	152
Interview duration	64.32	20.05	152
Height (cm), wave 1	163.05	5.65	152
Language ability, wave 1	3.66	0.61	152
German citizenship, wave 1	0.34		152
Religiosity, wave 1	1.65	1.33	152
Other non-German background			
Log hourly earnings, trimmed 5% / 95% quant.	2.35	0.49	521
Attractiveness rating, wave 1, bin.	0.40		521
Attractiveness rating, wave 1, cont.	5.90	1.24	521
Age	34.95	7.68	521

	mean	sd	count
Sample: Cohort 1	0.16		521
Sample: Cohort 2	0.29		521
Sample: Cohort 3	0.55		521
Currently enrolled	0.07		521
No degree/lower secondary	0.09		521
Upper secondary, vocational	0.26		521
Upper sec. general, post sec.	0.17		521
Tertiary	0.40		521
Parent. No degree/lower secondary	0.19		521
Parent. Upper secondary, vocational	0.43		521
Parent. Upper sec. general, post sec.	0.07		521
Parent. Tertiary	0.31		521
B5 - extraversion, wave 1, factor score	-0.16	0.83	521
B5 - agreeableness, wave 1, factor score	-0.33	0.83	521
B5 - conscientiousness, wave 1, factor score	0.27	0.74	521
B5 - neuroticism, wave 1, factor score	0.01	0.81	521
B5 - openness, wave 1, factor score	0.15	0.87	521
Self-rated health	3.61	1.00	521
BMI, wave 1	23.03	4.00	521
Interview duration	64.92	21.05	521
Height (cm), wave 1	166.04	6.22	521
Language ability, wave 1	3.69	0.58	521
German citizenship, wave 1	0.32		521
Religiosity, wave 1	1.48	1.37	521

TABLE A.2: Men: sample characteristics for preferred specification (N = 7,507).

	mean	sd	count
Native German			
Log hourly earnings, trimmed 5% / 95% quant.	2.53	0.53	6217
Attractiveness rating, wave 1, bin.	0.20		6217
Attractiveness rating, wave 1, cont.	5.33	1.32	6217
Age	32.24	7.93	6217
Sample: Cohort 1	0.26		6217
Sample: Cohort 2	0.35		6217
Sample: Cohort 3	0.39		6217
Currently enrolled	0.12		6217
No degree/lower secondary	0.04		6217
Upper secondary, vocational	0.39		6217
Upper sec. general, post sec.	0.12		6217
Tertiary	0.33		6217

Parent. No degree/lower secondary	0.05		6217
Parent. Upper secondary, vocational	0.59		6217
Parent. Upper sec. general, post sec.	0.05		6217
Parent. Tertiary	0.31		6217
B5 - extraversion, wave 1, factor score	0.06	0.87	6217
B5 - agreeableness, wave 1, factor score	0.03	0.80	6217
B5 - conscientiousness, wave 1, factor score	0.02	0.79	6217
B5 - neuroticism, wave 1, factor score	-0.28	0.76	6217
B5 - openness, wave 1, factor score	-0.03	0.75	6217
Self-rated health	3.80	0.91	6217
BMI, wave 1	24.88	4.39	6217
Interview duration	57.19	19.27	6217
Height (cm), wave 1	181.22	7.05	6217
Language ability, wave 1	3.98	0.19	6217
German citizenship, wave 1	1.00		6217
Religiosity, wave 1	1.52	1.43	6217
Ethnic-German Immigrant (Aussiedler)			
Log hourly earnings, trimmed 5% / 95% quant.	2.52	0.50	339
Attractiveness rating, wave 1, bin.	0.20		339
Attractiveness rating, wave 1, cont.	5.35	1.27	339
Age	30.42	6.85	339
Sample: Cohort 1	0.27		339
Sample: Cohort 2	0.49		339
Sample: Cohort 3	0.24		339
Currently enrolled	0.13		339
No degree/lower secondary	0.08		339
Upper secondary, vocational	0.47		339
Upper sec. general, post sec.	0.14		339
Tertiary	0.17		339
Parent. No degree/lower secondary	0.04		339
Parent. Upper secondary, vocational	0.55		339
Parent. Upper sec. general, post sec.	0.08		339
Parent. Tertiary	0.33		339
B5 - extraversion, wave 1, factor score	0.08	0.78	339
B5 - agreeableness, wave 1, factor score	-0.01	0.69	339
B5 - conscientiousness, wave 1, factor score	0.01	0.70	339
B5 - neuroticism, wave 1, factor score	-0.04	0.75	339
B5 - openness, wave 1, factor score	-0.14	0.75	339
Self-rated health	3.86	0.92	339
BMI, wave 1	24.89	4.23	339
Interview duration	63.74	20.48	339

Height (cm), wave 1	180.56	6.69	339
Language ability, wave 1	3.73	0.47	339
German citizenship, wave 1	1.00		339
Religiosity, wave 1	1.52	1.38	339
Multiethnic ("Half-German")			
Log hourly earnings, trimmed 5% / 95% quant.	2.57	0.55	351
Attractiveness rating, wave 1, bin.	0.25		351
Attractiveness rating, wave 1, cont.	5.42	1.32	351
Age	32.40	7.78	351
Sample: Cohort 1	0.25		351
Sample: Cohort 2	0.34		351
Sample: Cohort 3	0.40		351
Currently enrolled	0.12		351
No degree/lower secondary	0.02		351
Upper secondary, vocational	0.40		351
Upper sec. general, post sec.	0.11		351
Tertiary	0.36		351
Parent. No degree/lower secondary	0.11		351
Parent. Upper secondary, vocational	0.54		351
Parent. Upper sec. general, post sec.	0.06		351
Parent. Tertiary	0.29		351
B5 - extraversion, wave 1, factor score	0.12	0.85	351
B5 - agreeableness, wave 1, factor score	0.08	0.83	351
B5 - conscientiousness, wave 1, factor score	0.00	0.82	351
B5 - neuroticism, wave 1, factor score	-0.29	0.82	351
B5 - openness, wave 1, factor score	-0.02	0.74	351
Self-rated health	3.69	0.97	351
BMI, wave 1	24.97	4.21	351
Interview duration	60.63	18.81	351
Height (cm), wave 1	181.01	7.80	351
Language ability, wave 1	3.99	0.17	351
German citizenship, wave 1	0.96		351
Religiosity, wave 1	1.35	1.44	351
Turkish background			
Log hourly earnings, trimmed 5% / 95% quant.	2.55	0.54	220
Attractiveness rating, wave 1, bin.	0.24		220
Attractiveness rating, wave 1, cont.	5.65	1.11	220
Age	32.92	8.42	220
Sample: Cohort 1	0.27		220
Sample: Cohort 2	0.22		220
Sample: Cohort 3	0.50		220

Currently enrolled	0.15		220
No degree/lower secondary	0.21		220
Upper secondary, vocational	0.36		220
Upper sec. general, post sec.	0.08		220
Tertiary	0.20		220
Parent. No degree/lower secondary	0.50		220
Parent. Upper secondary, vocational	0.39		220
Parent. Upper sec. general, post sec.	0.03		220
Parent. Tertiary	0.08		220
B5 - extraversion, wave 1, factor score	0.38	0.83	220
B5 - agreeableness, wave 1, factor score	0.02	0.70	220
B5 - conscientiousness, wave 1, factor score	0.01	0.80	220
B5 - neuroticism, wave 1, factor score	-0.03	0.80	220
B5 - openness, wave 1, factor score	-0.20	0.69	220
Self-rated health	3.70	0.91	220
BMI, wave 1	27.80	21.65	220
Interview duration	62.61	22.11	220
Height (cm), wave 1	173.23	15.60	220
Language ability, wave 1	3.65	0.53	220
German citizenship, wave 1	0.38		220
Religiosity, wave 1	1.62	1.21	220
Other non-German background			
Log hourly earnings, trimmed 5% / 95% quant.	2.46	0.56	380
Attractiveness rating, wave 1, bin.	0.20		380
Attractiveness rating, wave 1, cont.	5.30	1.31	380
Age	32.45	7.68	380
Sample: Cohort 1	0.23		380
Sample: Cohort 2	0.37		380
Sample: Cohort 3	0.40		380
Currently enrolled	0.11		380
No degree/lower secondary	0.18		380
Upper secondary, vocational	0.28		380
Upper sec. general, post sec.	0.14		380
Tertiary	0.29		380
Parent. No degree/lower secondary	0.27		380
Parent. Upper secondary, vocational	0.29		380
Parent. Upper sec. general, post sec.	0.10		380
Parent. Tertiary	0.34		380
B5 - extraversion, wave 1, factor score	0.14	0.67	380
B5 - agreeableness, wave 1, factor score	-0.09	0.76	380
B5 - conscientiousness, wave 1, factor score	-0.09	0.87	380

B5 - neuroticism, wave 1, factor score	-0.05	0.77	380
B5 - openness, wave 1, factor score	-0.21	0.83	380
Self-rated health	3.76	0.93	380
BMI, wave 1	25.47	4.83	380
Interview duration	59.29	19.83	380
Height (cm), wave 1	177.19	7.32	380
Language ability, wave 1	3.62	0.57	380
German citizenship, wave 1	0.40		380
Religiosity, wave 1	1.50	1.33	380

A.2 Specification curves
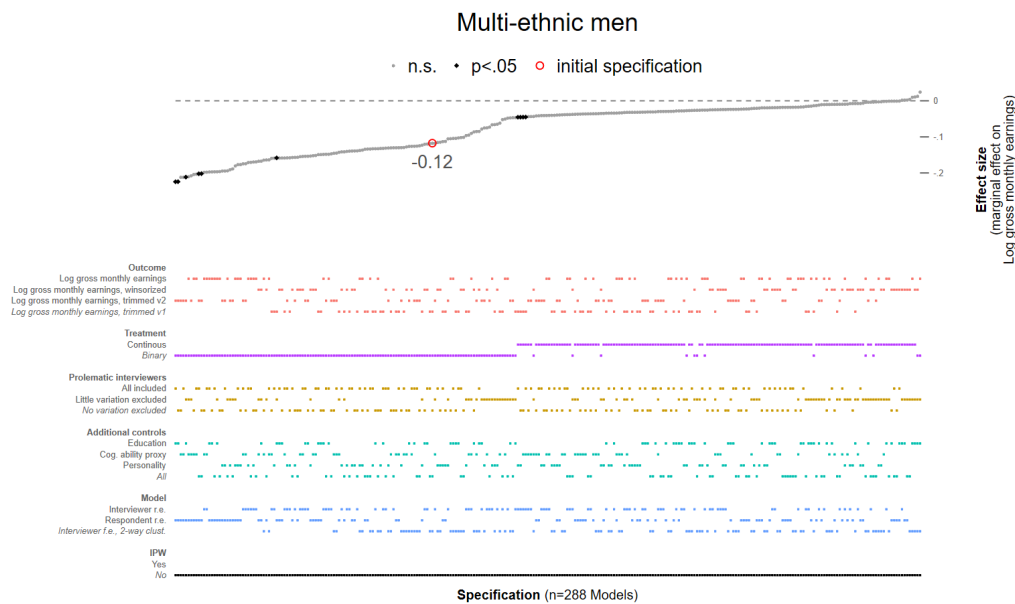
FIGURE A.2: Specification curve: multiethnic immigrant men. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.

Men with other non-German background

· n.s.   • p<.05   ○ initial specification
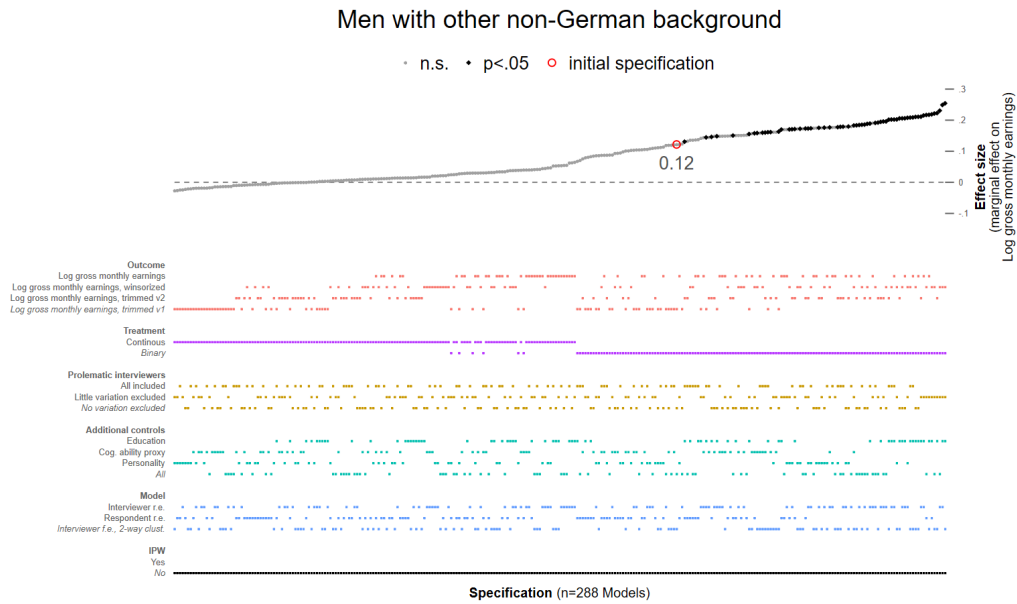
FIGURE A.3: Specification curve: immigrant men, "other" background. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.

Multi-ethnic women

· n.s.   • p<.05   ○ initial specification
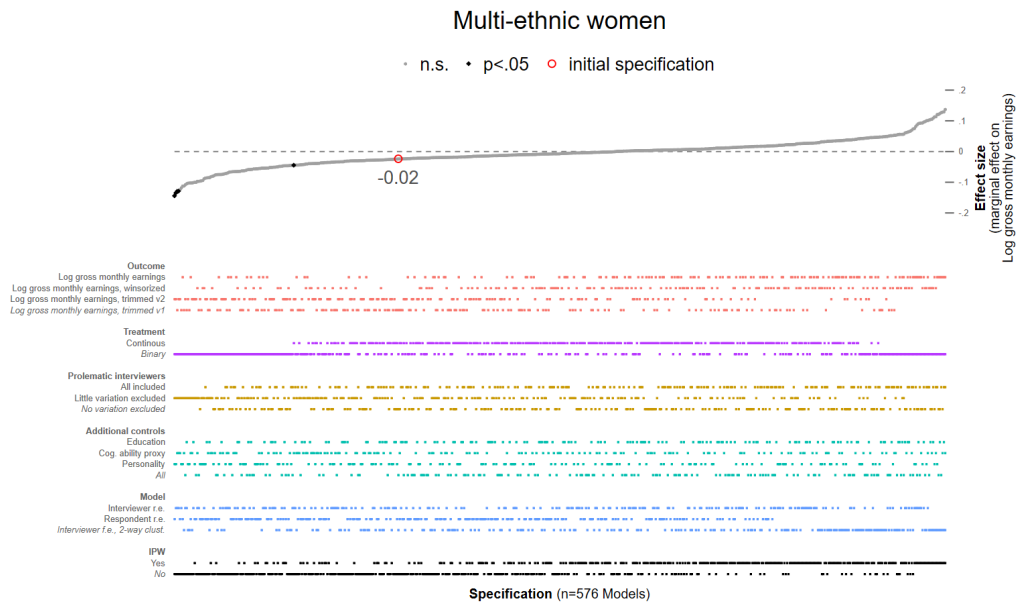
FIGURE A.4: Specification curve: multiethnic immigrant women. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.
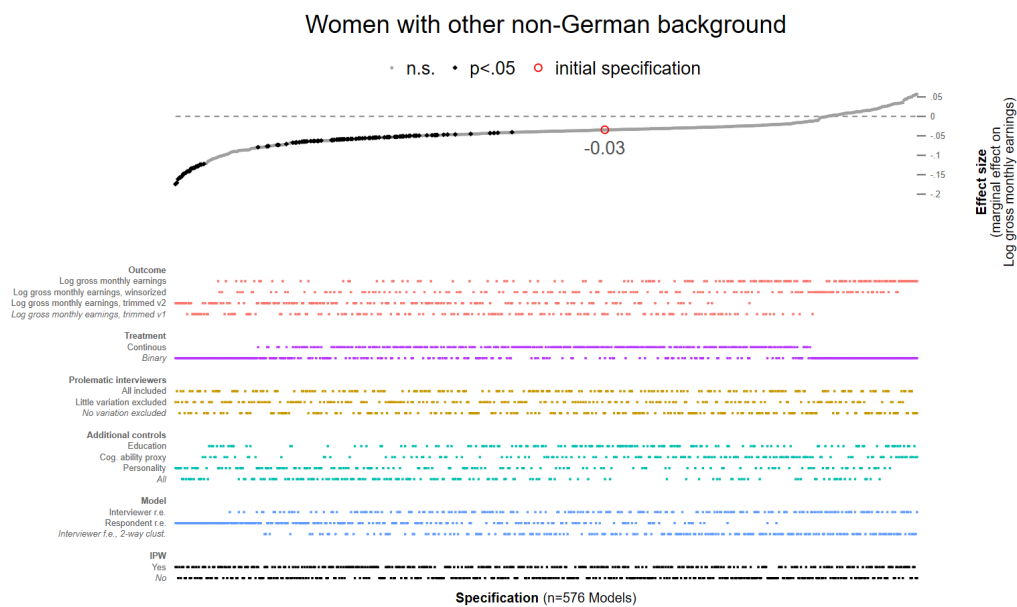
FIGURE A.5: Specification curve: immigrant women, "other" background. Each dot in the curve (above the dotted line) represents the size of the marginal effect of physical attractiveness on earnings. The dots below the dotted line are vertically arranged underneath each estimate to denote the analytical decisions that led to each result. n.s.: not significant.

Appendix B

Appendix to Chapter 3

B.1 Summary statistics

Ethnicity	Male applicants			Female applicants			Total
	Low attr.	Average attr.	High attr.	Low attr.	Average attr.	High attr.	
Ethnic majority German	79	83	124	88	98	92	*564*
German low prototypicality	122	45	117	58	113	107	*562*
Danish	95	97	103	94	85	70	*544*
Greek	92	84	94	82	79	93	*524*
Turkish low prototypicality	117	50	118	54	111	107	*557*
Turkish	190	218	162	77	91	90	*828*
Turkish religious	0	0	0	89	90	98	*277*
Total	*695*	*577*	*718*	*542*	*667*	*657*	*3,856*

TABLE B.1: Applications sent by ethnicity, gender and attractiveness level.

Response type	Count	Percentage
Positive response types	*1,626*	*42.2%*
Invitation	1,099	28.5%
Pre-invitation	257	6.7%
Request for information	270	7.0%
Unclear response types	*171*	*4.4%*
Missed call	25	0.0%
Application incomplete	90	2.3%
Other	56	1.5%
Negative response types	*2,059*	*53.4%*
Confirmation of receipt	191	5.0%
Rejection	820	21.3%
No response	1,048	27.2%
Total	*3,856*	*100.0%*

TABLE B.2: Responses classified by type, using the classification scheme from the GEMM study (Lancee et al., 2019).

B.2 Supplementary analyses

Comparison	Prop. 1	Prop. 2	z	p
Migration background (1) vs. no migration background (2)	0.409	0.520	-6.124	0.000 ***
Low or average attractiveness (1) vs. high attractiveness (2)	0.426	0.469	-2.500	0.012 *
Men only: low or average attractiveness (1) vs. high attractiveness (2)	0.389	0.438	-2.085	0.037 *
Women only: low or average attractiveness (1) vs. high attractiveness (2)	0.465	0.502	-1.521	0.128
Average or high attractiveness (1) vs. low attractiveness (2)	0.462	0.397	3.736	0.000 ***
Men only: average or high attractiveness (1) vs. low attractiveness (2)	0.424	0.373	2.165	0.030 *
Women only: average or high attractiveness (1) vs. low attractiveness (2)	0.499	0.427	2.758	0.006 **
Turkish (1) vs. German (2)	0.362	0.520	-8.060	0.000 ***

Greek (1) vs. German (2)	0.437	0.520	-3.049	0.002 **
Danish (1) vs. German (2)	0.526	0.520	0.227	0.820
Male (1) vs. female (2)	0.406	0.478	-4.370	0.000 ***
Turkish men (1) vs. German men (2)	0.326	0.494	-6.201	0.000 ***
Greek men (1) vs. German men (2)	0.392	0.494	-2.677	0.007 **
Danish men (1) vs. German men (2)	0.486	0.494	-0.210	0.834
Turkish women (1) vs. German women (2)	0.400	0.546	-5.200	0.000 ***
Greek women (1) vs. German women (2)	0.484	0.546	-1.628	0.104
Danish women (1) vs. German women (2)	0.572	0.546	0.662	0.508
Low-attractiveness German men (1) vs. high-attractiveness German men (2)	0.440	0.530	-1.845	0.065
Low-attractiveness Turkish men (1) vs. high-attractiveness Turkish men (2)	0.315	0.322	-0.174	0.862
Low-attractiveness German women (1) vs. high-attractiveness German women (2)	0.507	0.630	-2.224	0.026 *
Low-attractiveness Turkish women (1) vs. high-attractiveness Turkish women (2)	0.332	0.428	-2.163	0.031 *
Low customer contact (1) vs. high customer contact (2)	0.429	0.460	-1.869	0.062
Low-contact occupations: Low attractiveness (1) vs. high attractiveness (2)	0.376	0.478	-4.018	0.000 ***
High-contact occupations: Low attractiveness (1) vs. high attractiveness (2)	0.431	0.455	-0.747	0.455
No or average references (1) vs. excellent references (2)	0.417	0.490	-4.214	0.000 ***
Average or excellent references (1) vs. no references (2)	0.456	0.411	2.584	0.010 **
Turkish women without (1) vs. with headscarf (2)	0.470	0.271	5.410	0.000 ***

TABLE B.3: Two-sample tests of proportions for callback rates between groups of interest. * $p < .05$, ** $p < .01$, *** $p < .001$.

	(1)
Attractiveness (ref. low)	
Average attractiveness	1.156
	(0.137)
High attractiveness	1.307 *
	(0.145)
Female applicant (ref. male)	1.252
	(0.149)
Attractiveness x gender (ref. low x male)	
Average attractiveness # Female	1.139
	(0.191)
High attractiveness # Female	1.036
	(0.169)
Intercept	0.596 **
	(0.048)
Number of observations	3685

TABLE B.4: Effects of gender and attractiveness on callback rates. Logistic regression results presented as odds ratios. Standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.
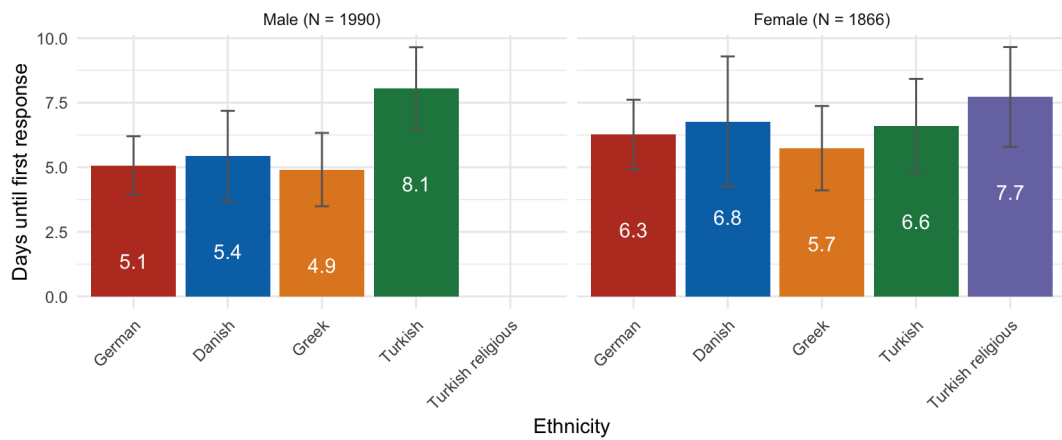
FIGURE B.1: Mean number of days until first employer response by ethnicity and gender. $N = 3{,}856$.

	(1) Male applicants	(2) Female applicants	(3) Male applicants	(4) Female applicants
Ethnicity (ref. German majority)				
Danish	1.002	0.873		
	(0.255)	(0.234)		
Greek	0.681	0.923		
	(0.183)	(0.261)		
Turkish	0.585 **	0.736		
	(0.112)	(0.183)		
Attractiveness (ref. low)				
Average attractiveness	1.314	0.952		
	(0.308)	(0.210)		
High attractiveness	1.435	1.652 *		
	(0.282)	(0.374)		
Ethnicity x Attractiveness (ref. German x low attractiveness)				
Danish #			1.002	0.873
Low attractiveness			(0.255)	(0.234)
Danish #	0.944	2.608 *	1.243	2.168 **
Average attractiveness	(0.356)	(1.002)	(0.317)	(0.629)
Danish #	0.959	0.860	1.379	1.240
High attractiveness	(0.336)	(0.340)	(0.343)	(0.372)
German #			1.314	0.952
Average attractiveness			(0.308)	(0.210)
German #			1.435	1.652 *
High attractiveness			(0.282)	(0.374)
Greek #			0.681	0.923
Low attractiveness			(0.183)	(0.261)
Greek #	0.835	1.326	0.747	1.166
Average attractiveness	(0.333)	(0.518)	(0.203)	(0.332)
Greek #	1.084	0.478	1.059	0.729
High attractiveness	(0.399)	(0.183)	(0.274)	(0.198)
Turkish #			0.585 **	0.736
Low attractiveness			(0.112)	(0.183)
Turkish #	0.856	1.345	0.658 *	0.942
Average attractiveness	(0.254)	(0.430)	(0.129)	(0.209)
Turkish #	0.719	0.718	0.604 **	0.873
High attractiveness	(0.192)	(0.234)	(0.118)	(0.195)
Intercept	0.787	1.029	0.787	1.029
	(0.114)	(0.174)	(0.114)	(0.174)
Number of observations	1892	1520	1892	1520

TABLE B.5: Effects of ethnicity and attractiveness on callback rates, including Danish applicants. Logistic regression results presented as odds ratios. Models 1 and 3 only male respondents; models 2 and 4 only female respondents, excluding veiled Turkish applicants. Standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.

	(1) All applicants		(2) Male applicants		(3) Female applicants	
Attractiveness (ref. low)						
Average attractiveness	1.322		1.196		1.421	
	(0.195)		(0.251)		(0.298)	
High attractiveness	1.571	**	1.542	*	1.592	*
	(0.229)		(0.310)		(0.337)	
Reference quality (ref. none)						
Average	1.063		1.121		0.994	
	(0.158)		(0.223)		(0.222)	
High	1.665	**	1.283		2.335	**
	(0.245)		(0.255)		(0.517)	
Attractiveness x reference quality **(ref. low attractiveness x no reference)**						
Average attractiveness # Average reference	1.102		1.049		1.184	
	(0.226)		(0.304)		(0.350)	
Average attractiveness # Excellent reference	0.827		0.870		0.691	
	(0.171)		(0.257)		(0.204)	
High attractiveness # Average reference	0.892		0.765		1.075	
	(0.180)		(0.211)		(0.320)	
High attractiveness # Excellent reference	0.707		0.811		0.580	
	(0.142)		(0.224)		(0.171)	
Intercept	0.541	**	0.526	**	0.560	**
	(0.058)		(0.076)		(0.089)	
Number of observations	3685		1892		1793	

TABLE B.6: Effects of attractiveness and reference quality on callback rates. Logistic regression results presented as odds ratios. Standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.
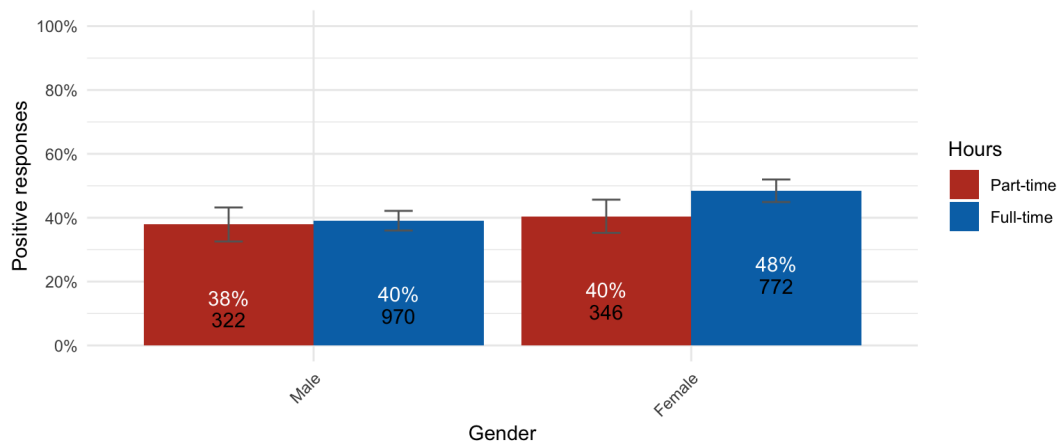
FIGURE B.2: Callback rates by gender and hours. N = 2,410 job postings, excluding those with missing or inconclusive information about hours. Total N for each bar shown in black.

	(1) All applicants		(2) Male applicants		(3) Female applicants	
Attractiveness (ref. low)						
Average attractiveness	1.258		1.234		1.171	
	(0.149)		(0.187)		(0.228)	
High attractiveness	1.528	**	1.467	**	1.578	*
	(0.170)		(0.200)		(0.306)	
Customer contact (ref. low)						
High contact	1.341	*	1.210		1.212	
	(0.174)		(0.232)		(0.235)	
Attractiveness x customer contact						
(ref. low attractiveness x low contact)						
Average attractiveness #	1.030		0.872		1.118	
High contact	(0.189)		(0.257)		(0.290)	
High attractiveness #	0.658	*	0.629		0.662	
High contact	(0.117)		(0.167)		(0.173)	
Intercept	0.647	**	0.585	**	0.811	
	(0.052)		(0.057)		(0.118)	
Number of observations	3153		1633		1520	

TABLE B.7: Effects of attractiveness and level of occupational customer contact on callback rates. Logistic regression results presented as odds ratios. Standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.
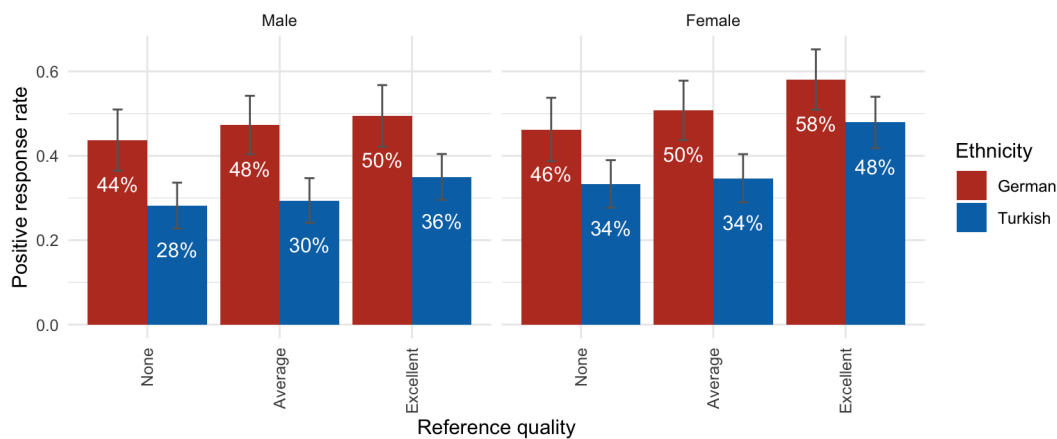
FIGURE B.3: Response rates by reference quality, gender, and ethnicity. Limited to only German and Turkish applicants, including ambiguous and religious conditions. $N = 2,788$.

	(1) All Turkish-background applicants		(2) Male Turkish-background applicants		(3) Female Turkish-background applicants	
Religious badge (ref. no badge)	0.517	**	0.748		0.359	**
	(0.124)		(0.244)		(0.131)	
Attractiveness (ref. low)						
Average attractiveness	0.951		0.807		1.067	
	(0.208)		(0.245)		(0.347)	
High attractiveness	1.005		0.659		1.402	
	(0.218)		(0.203)		(0.453)	
Religious badges x attractiveness (ref. no badge x low attractiveness)						
Religious badge #	1.540		1.528		1.467	
Average attractiveness	(0.498)		(0.678)		(0.713)	
Religious badge #	1.732		2.133		1.536	
High attractiveness	(0.569)		(1.010)		(0.726)	
Intercept	0.623	**	0.597	**	0.667	
	(0.094)		(0.115)		(0.163)	
Number of observations	1063		542		521	

TABLE B.8: Effects of attractiveness and religious badges on callback rates, only applicants with Turkish migration background. Logistic regression results presented as odds ratios. Standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.
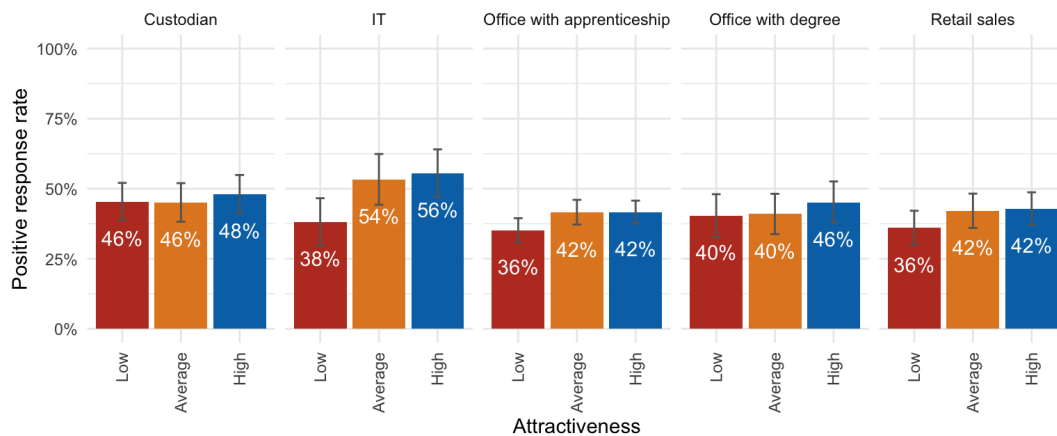
FIGURE B.4: Callback rates by physical attractiveness treatment and occupational grouping. "Office with apprenticeship" includes receptionist, secretary, and industrial clerk positions. "Office with degree" includes sales assistant and purchasing agent positions. $N = 3,856$.

Appendix C

Appendix to Chapter 4

C.1 Descriptive statistics

	Selected	Unselected	Diff	pvalue
Male	0.52	0.53	0.01	0.74
Born 1970 or later	0.44	0.42	-0.02	0.25
Abitur	0.41	0.38	-0.03	0.04
In work or training	0.66	0.66	0.01	0.72
Resident of East Germany	0.21	0.20	-0.01	0.55
Living with partner	0.61	0.62	0.01	0.65
N	1794	1568		

TABLE C.1: Means of demographic variables in the analytical sample ("Selected") and the rest of the sample ("Unselected") and t-test results.

C.2   Supplemental analyses: main results

	(1)	(2)	(3)	(4)	(5)
	Model 1	Model 2	Model 3	Model 4	Model 5
High attractiveness	0.08*				
	(0.04)				
MENA phenotype x Mehmet		0.29***			
		(0.05)			
Ambig. phenotype x Mehmet			0.16**		
			(0.05)		
High att. x Jonas				0.05	
				(0.05)	
High att. x Mehmet					0.11*
					(0.05)
Constant	2.69***	2.56***	2.67***	2.59***	2.79***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
N	1794	895	899	900	894

TABLE C.2: Full OLS regression results for main analyses without control variables, corresponding to results shown in Figure 4.2. Model 2 restricted to only profiles in unambiguous phenotype condition, Model 3 to only profiles with ambiguous phenotype. Model 4 restricted to only profiles in German ethnic condition, Model 5 to profiles in Turkish ethnic condition. Robust standard errors in parentheses, $^{+}p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

	(1)	(2)	(3)	(4)	(5)
	Model 1	Model 2	Model 3	Model 4	Model 5
High attractiveness	0.08*				
	(0.04)				
MENA phenotype x Mehmet		0.26***			
		(0.05)			
Ambig. phenotype x Mehmet			0.15**		
			(0.05)		
High att. x Jonas				0.04	
				(0.05)	
High att. x Mehmet					0.11*
					(0.05)
Abitur	0.13***	0.16**	0.09$^+$	0.05	0.20***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Male	-0.12**	-0.08$^+$	-0.13**	-0.05	-0.17***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Born 1970 or later	-0.17***	-0.13*	-0.22***	-0.16*	-0.19**
	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
In work or training	-0.06	-0.10	0.01	-0.02	-0.10
	(0.05)	(0.06)	(0.07)	(0.07)	(0.07)
Lives in East Germany	-0.11*	-0.06	-0.17**	-0.10	-0.12$^+$
	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
Living with partner	0.15***	0.15**	0.12*	0.17***	0.10$^+$
	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)
Constant	2.75***	2.60***	2.75***	2.60***	2.89***
	(0.05)	(0.08)	(0.07)	(0.07)	(0.08)
N	1794	895	899	900	894

TABLE C.3: Full OLS regression results for main analyses including control variables, corresponding to results shown in Figure C.1. Model 2 restricted to only profiles in unambiguous phenotype condition, Model 3 to only profiles with ambiguous phenotype. Model 4 restricted to only profiles in German ethnic condition, Model 5 to profiles in Turkish ethnic condition. Robust standard errors in parentheses, $^+p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.
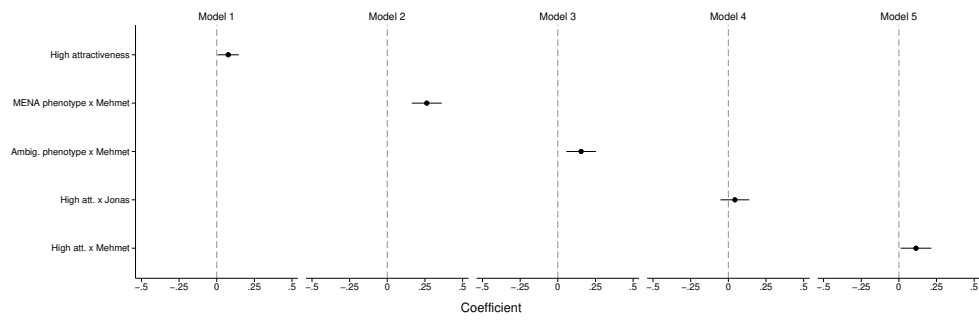
FIGURE C.1: Coefficients and 95% confidence intervals from multivariate regression analysis (OLS) with control variables. Full sample, N = 1,794.

	(1) Model 1	(2) Model 2
High attractiveness	0.05 (0.05)	0.05 (0.05)
Mehmet	0.20*** (0.05)	0.18*** (0.05)
High att. x Mehmet	0.06 (0.07)	0.06 (0.07)
Abitur		0.12*** (0.04)
Male		-0.11** (0.04)
Born 1970 or later		-0.17*** (0.04)
In work or training		-0.05 (0.05)
Lives in East Germany		-0.11* (0.04)
Living with partner		0.13*** (0.04)
Constant	2.59*** (0.03)	2.66*** (0.06)
N	1794	1794

TABLE C.4: Full OLS regression results testing the size of the beauty premium in trustworthiness across ethnic groups, with and without control variables. Robust standard errors in parentheses, $^+p < .1$. $^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

| | (1) | (2) | (3) | (4) | (5) |
	Model 1	Model 2	Model 3	Model 4	Model 5
High attractiveness	0.20*				
	(0.09)				
MENA phenotype x Mehmet		0.65***			
		(0.13)			
Ambig. phenotype x Mehmet			0.42***		
			(0.13)		
High att. x Jonas				0.11	
				(0.13)	
High att. x Mehmet					0.29*
					(0.13)
Abitur	0.31***	0.37**	0.19	0.12	0.44***
	(0.09)	(0.13)	(0.13)	(0.13)	(0.12)
Male	-0.28**	-0.21	-0.32*	-0.14	-0.41**
	(0.09)	(0.13)	(0.13)	(0.13)	(0.13)
Born 1970 or later	-0.45***	-0.34*	-0.60***	-0.40*	-0.48**
	(0.11)	(0.15)	(0.17)	(0.17)	(0.16)
In work or training	-0.16	-0.26	0.03	-0.08	-0.23
	(0.12)	(0.16)	(0.17)	(0.17)	(0.16)
Lives in East Germany	-0.29**	-0.17	-0.41*	-0.28$^+$	-0.29$^+$
	(0.11)	(0.16)	(0.16)	(0.16)	(0.16)
Living with partner	0.35***	0.35*	0.33*	0.45***	0.23$^+$
	(0.09)	(0.14)	(0.13)	(0.13)	(0.14)
cut1	-5.44***	-5.32***	-5.29***	-5.77***	-5.25***
	(0.36)	(0.54)	(0.49)	(0.61)	(0.46)
cut2	-3.64***	-3.34***	-3.66***	-3.34***	-4.02***
	(0.19)	(0.28)	(0.26)	(0.25)	(0.29)
cut3	-0.50***	-0.17	-0.50**	-0.19	-0.84***
	(0.13)	(0.20)	(0.18)	(0.18)	(0.20)
cut4	1.66***	1.98***	1.72***	2.15***	1.25***
	(0.14)	(0.20)	(0.19)	(0.20)	(0.20)
N	1794	895	899	900	894

TABLE C.5: Full ordered logistic regression results for main analyses including control variables. Model 2 restricted to only profiles in unambiguous phenotype condition, Model 3 to only profiles with ambiguous phenotype. Model 4 restricted to only profiles in German ethnic condition, Model 5 to profiles in Turkish ethnic condition. Robust standard errors in parentheses, $^+p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
Pre-rated attractiveness	-0.01		0.01	0.01
	(0.03)		(0.04)	(0.04)
Mehmet			0.09	0.03
			(0.36)	(0.35)
Pre-rated att. x Jonas		-0.00		
		(0.03)		
Pre-rated att. x Mehmet		0.05	0.04	0.04
		(0.03)	(0.06)	(0.06)
Abitur				0.16**
				(0.05)
Male				-0.08
				(0.05)
Born 1970 or later				-0.13*
				(0.06)
In work or training				-0.10
				(0.07)
Living with partner				0.15**
				(0.05)
Lives in East Germany				-0.05
				(0.06)
Constant	2.77***	2.56***	2.52***	2.57***
	(0.18)	(0.18)	(0.23)	(0.23)
N	895	895	895	895

TABLE C.6: Full OLS regression results testing the size of the beauty premium in trustworthiness across ethnic groups using continuous measures of facial attractiveness from pre-test ratings. Analyses limited to unambiguous faces (i.e., those used with only one name) to more strictly test the effects of appearance rather than name. Robust standard errors in parentheses, $^+p < .1$. $^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5
High attractiveness	0.08$^+$ (0.04)				
MENA phenotype x Mehmet		0.25*** (0.06)			
Ambig. phenotype x Mehmet			0.13* (0.06)		
High att. x Jonas				0.01 (0.06)	
High att. x Mehmet					0.13* (0.06)
Abitur	0.14*** (0.04)	0.17** (0.06)	0.10$^+$ (0.06)	0.06 (0.06)	0.22*** (0.06)
Male	-0.16*** (0.04)	-0.13* (0.06)	-0.17** (0.06)	-0.13* (0.06)	-0.19** (0.06)
Born 1970 or later	-0.14** (0.05)	-0.06 (0.07)	-0.23** (0.07)	-0.16* (0.07)	-0.10 (0.07)
In work or training	-0.06 (0.05)	-0.12 (0.07)	0.01 (0.07)	-0.02 (0.07)	-0.11 (0.07)
Lives in East Germany	-0.13* (0.06)	-0.03 (0.08)	-0.23** (0.08)	-0.12 (0.07)	-0.14$^+$ (0.08)
Living with partner	0.11* (0.04)	0.13* (0.06)	0.09 (0.06)	0.11$^+$ (0.06)	0.10 (0.07)
Lives in urban area	-0.09$^+$ (0.05)	-0.16* (0.08)	-0.03 (0.07)	-0.09 (0.07)	-0.10 (0.08)
Pct. migrants in community	-0.00 (0.00)	0.00 (0.01)	-0.01 (0.01)	-0.01$^+$ (0.01)	0.01 (0.01)
Constant	2.89*** (0.07)	2.73*** (0.11)	2.91*** (0.10)	2.88*** (0.11)	2.91*** (0.11)
N	1370	677	693	671	699

TABLE C.7: Full OLS regression results including controls related to respondents' location of residence, for respondents with valid postal code. Model 2 restricted to only profiles in unambiguous phenotype condition, Model 3 to only profiles with ambiguous phenotype. Model 4 restricted to only profiles in German ethnic condition, Model 5 to profiles in Turkish ethnic condition. "Urban area" defined as a community of 10,000 residents or more. Percentage of migrants and community population at the postal code level derived from 2011 Zensus data. Robust standard errors in parentheses, $^+p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

C.3 Supplemental analyses: heterogeneous effects

C.3.1 Immigration attitudes

	(1)	(2)	(3)	(4)
	Model 1a	Model 1b	Model 2a	Model 2b
Jonas x inclusionary	0.00	0.16**		
	(.)	(0.06)		
Jonas x exclusionary	-0.16**	0.00		
	(0.06)	(.)		
Mehmet x inclusionary	0.26***	0.43***		
	(0.04)	(0.06)		
Mehmet x exclusionary	-0.21**	-0.04		
	(0.08)	(0.09)		
Low att. x Mehmet x inclusionary (ref.)			0.00	
			(.)	
High att. x Mehmet x inclusionary			0.12*	
			(0.06)	
Low att. x Mehmet x exclusionary (ref.)				0.00
				(.)
High att. x Mehmet x exclusionary				0.03
				(0.14)
Constant	2.67***	2.51***	2.87***	2.45***
	(0.03)	(0.05)	(0.04)	(0.10)
N	1638	1638	652	160

TABLE C.8: Full OLS regression results for analyses divided by immigration attitudes, corresponding to results shown in Figure 3. Robust standard errors in parentheses, $^{+}p < .1$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

C.3.2   Mosque support

	(1)	(2)	(3)	(4)
	Model 1a	Model 1b	Model 2a	Model 2b
Jonas x pro-mosque	0.00	0.03		
	(.)	(0.05)		
Jonas x anti-mosque	-0.03	0.00		
	(0.05)	(.)		
Mehmet x pro-mosque	0.36***	0.40***		
	(0.05)	(0.05)		
Mehmet x anti-mosque	0.03	0.07		
	(0.06)	(0.06)		
Low att. x Mehmet x pro-mosque (ref.)			0.00	
			(.)	
High att. x Mehmet x pro-mosque			0.11	
			(0.07)	
Low att. x Mehmet x anti-mosque (ref.)				0.00
				(.)
High att. x Mehmet x anti-mosque				0.04
				(0.09)
Constant	2.66***	2.62***	2.96***	2.67***
	(0.04)	(0.04)	(0.05)	(0.05)
N	1540	1540	369	395

TABLE C.9: Full OLS regression results for analyses divided by support or opposition to mosque construction. Robust standard errors in parentheses, $^{+}p < .1$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

	(1)	(2)	(3)	(4)
	Model 1	Model 2	Model 3	Model 4
Mehmet	0.26***	0.25***	0.36***	0.35***
	(0.04)	(0.04)	(0.05)	(0.05)
Exclusionary attitudes	-0.16**	-0.13*		
	(0.06)	(0.06)		
Exclusionary x Mehmet	-0.30**	-0.29**		
	(0.10)	(0.10)		
Opposes mosque			-0.03	-0.03
			(0.05)	(0.05)
Opposes mosque x Mehmet			-0.29***	-0.29***
			(0.08)	(0.08)
Male		-0.10**		-0.15***
		(0.04)		(0.04)
Born 1970 or later		-0.18***		-0.16**
		(0.05)		(0.05)
Abitur		0.05		0.09*
		(0.04)		(0.04)
In work or training		-0.00		-0.04
		(0.05)		(0.05)
Lives in East Germany		-0.08+		-0.10*
		(0.05)		(0.05)
Living with partner		0.11**		0.13**
		(0.04)		(0.04)
Constant	2.67***	2.73***	2.66***	2.74***
	(0.03)	(0.06)	(0.04)	(0.06)
N	1638	1638	1540	1540

TABLE C.10: Full OLS regression results testing the extent of ethnic bias in trustworthiness perceptions across groups defined by immigration attitudes and attitudes toward Islam, with and without control variables. Robust standard errors in parentheses, $^+p < .1$. $^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

C.3.3 Respondent gender
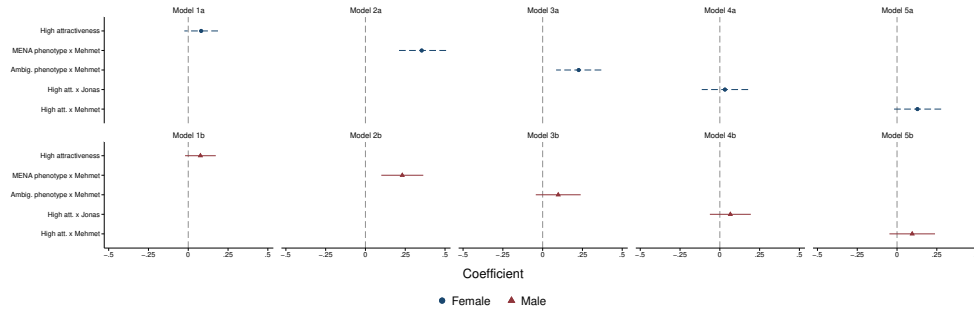
FIGURE C.2: Coefficients and 95% confidence intervals from multi-variate regression analysis (OLS) without control variables. Sample divided into male (n = 937) and female respondents (n = 857).

	(1)	(2)	(3)	(4)	(5)
	Model 1a	Model 2a	Model 3a	Model 4a	Model 5a
High attractiveness	0.08				
	(0.05)				
MENA phenotype x Mehmet		0.35***			
		(0.08)			
Ambig. phenotype x Mehmet			0.23**		
			(0.07)		
High att. x Jonas				0.03	
				(0.07)	
High att. x Mehmet					0.13+
					(0.08)
Constant	2.74***	2.57***	2.69***	2.62***	2.86***
	(0.04)	(0.06)	(0.05)	(0.05)	(0.05)
N	857	408	449	425	432

TABLE C.11: Full OLS regression results for female respondents only, corresponding to results shown in Figure C.2. Robust standard errors in parentheses, $^+ p < .1$. $^* p < .05$. $^{**} p < .01$. $^{***} p < .001$.

	(1) Model 1b	(2) Model 2b	(3) Model 3b	(4) Model 4b	(5) Model 5b
High attractiveness	0.08				
	(0.05)				
MENA phenotype x Mehmet		0.23***			
		(0.07)			
Ambig. phenotype x Mehmet			0.10		
			(0.07)		
High att. x Jonas				0.07	
				(0.07)	
High att. x Mehmet					0.10
					(0.07)
Constant	2.64***	2.55***	2.65***	2.56***	2.72***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
N	937	487	450	475	462

TABLE C.12: Full OLS regression results for male respondents only, corresponding to results shown in Figure C.2. Robust standard errors in parentheses, $^+ p < .1$. $^* p < .05$. $^{**} p < .01$. $^{***} p < .001$.

C.3.4   Socioeconomic status

	(1) Model 1a	(2) Model 1b	(3) Model 2a	(4) Model 2b
Jonas x no Abitur	0.00	-0.01	0.00	-0.06
	(.)	(0.05)	(.)	(0.05)
Jonas x Abitur	0.01	0.00	0.06	0.00
	(0.05)	(.)	(0.05)	(.)
Mehmet x no Abitur	0.17***	0.16**	0.15**	0.10$^+$
	(0.05)	(0.05)	(0.05)	(0.05)
Mehmet x Abitur	0.31***	0.29***	0.34***	0.29***
	(0.05)	(0.05)	(0.05)	(0.05)
Male			-0.11**	-0.11**
			(0.04)	(0.04)
Born 1970 or later			-0.17***	-0.17***
			(0.04)	(0.04)
In work or training			-0.05	-0.05
			(0.05)	(0.05)
Lives in East Germany			-0.11**	-0.11**
			(0.04)	(0.04)
Living with partner			0.14***	0.14***
			(0.04)	(0.04)
Constant	2.61***	2.62***	2.71***	2.77***
	(0.03)	(0.04)	(0.05)	(0.06)
N	1794	1794	1794	1794

TABLE C.13: Full OLS regression results testing the extent of ethnic bias in trustworthiness perceptions across groups defined by educational attainment (having received or not received the *Abitur*, or German matriculation examination), with and without control variables. Robust standard errors in parentheses, $^+p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

	(1)	(2)	(3)	(4)
	Model 1a	Model 1b	Model 2a	Model 2b
Pre-rated social class	0.03	-0.10	0.04	-0.09
	(0.07)	(0.09)	(0.07)	(0.09)
Mehmet	0.18***	0.27***	0.16**	0.26***
	(0.05)	(0.06)	(0.05)	(0.06)
Male			-0.18***	-0.00
			(0.05)	(0.05)
Born 1970 or later			-0.19***	-0.14*
			(0.06)	(0.07)
In work or training			-0.11$^+$	0.07
			(0.06)	(0.07)
Lives in East Germany			-0.15**	-0.06
			(0.06)	(0.07)
Living with partner			0.13*	0.15**
			(0.05)	(0.06)
Constant	2.41***	3.23***	2.53***	3.15***
	(0.43)	(0.54)	(0.42)	(0.54)
N	1051	743	1051	743

TABLE C.14: Full OLS regression results for analyses divided by respondent educational attainment (having received or not received the *Abitur*, or German matriculation examination). Models 1a and 2a include only respondents without *Abitur*, models 1b and 2b include only respondents with *Abitur*. Robust standard errors in parentheses, $^+p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

Appendix D

Appendix to Chapter 5

D.1 Summary statistics

´

TABLE D.1: Respondent characteristics

	N	mean	min	max	sd
Age	2158	41.089	18	98	15.347
Male	2166	0.486	0	1	0.500
High school diploma or higher	2141	0.584	0	1	0.493
University diploma or higher	2141	0.249	0	1	0.433
Annual household income (TRY)	1962	35987.371	0	400000	27856.697
Self-rated social class	2152	4.523	0	10	2.003
Muslim	2142	0.965	0	1	0.185
Self-rated religiosity	2148	3.602	1	5	0.801
Wears veil (women only)	1099	0.564	0	1	0.496
Wears beard (men only)	1027	0.655	0	1	0.475
Voted for AKP in last election	1505	0.395	0	1	0.489
Lives in rural area	2170	0.127	0	1	0.333

D.2 Religiosity treatment: manipulation check
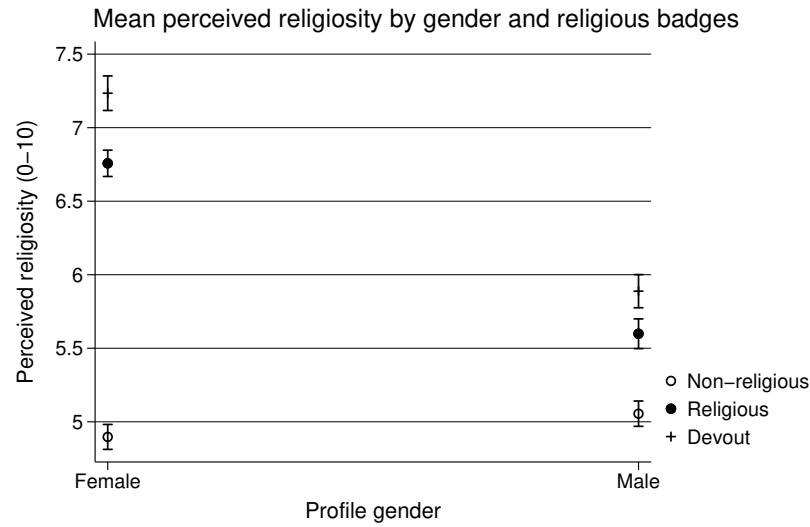
FIGURE D.1: Perceptions of religiosity by vignette person gender and religiosity.

Notes: Figure D.1 shows that women with headscarves both wearing a türban (religious) or a chador (devout) are perceived to be significantly more religious than without headcovering (average rating 7.23 (devout) vs 6.76 (religious) vs. 4.90 (non-religious) on a scale from 0 to 10). For men, bearded profiles in the devout (average rating 5.89) and religious (5.60) conditions are perceived to be more religious than non-bearded men (5.06). However, comparing the two different signals of religiosity (headscarves vs. beards), we see that head coverings act as a much stronger signal of religiosity than beards. $N = 12,782$ ratings from 2,133 respondents.

D.3 Supplementary analyses
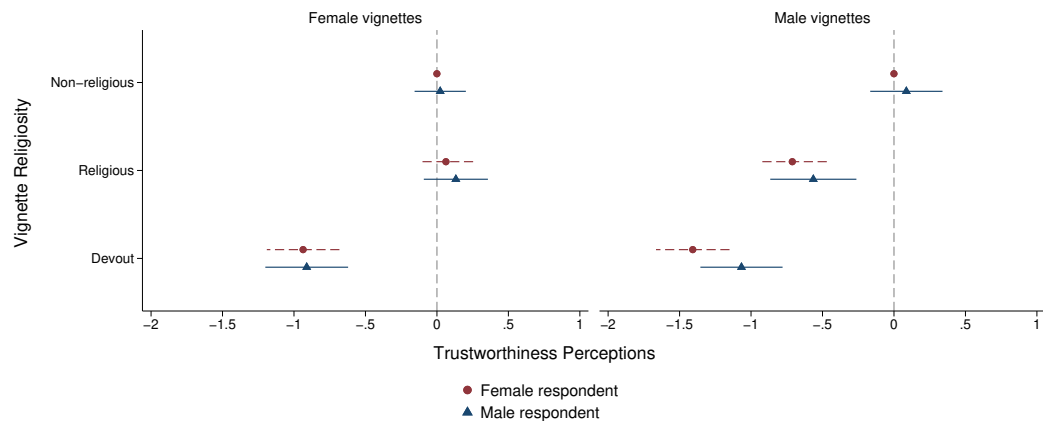
FIGURE D.2: Perceptions of trustworthiness by vignette religiosity and gender as well as respondent gender. $N = 12,893$ ratings from 2,158 respondents.

TABLE D.2: Main results of religiosity treatment on perceived trustworthiness by vignette gender

	Female vignettes				Male vignettes			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Non-religious vignette	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Religious vignette	0.09 (0.08)	0.08 (0.08)	0.09 (0.08)	0.09 (0.08)	-0.68*** (0.09)	-0.68*** (0.09)	-0.68*** (0.09)	-0.68*** (0.09)
Devout vignette	-0.94*** (0.11)	-0.94*** (0.11)	-0.94*** (0.11)	-0.94*** (0.11)	-1.28*** (0.10)	-1.28*** (0.10)	-1.29*** (0.10)	-1.28*** (0.09)
High social class		0.68*** (0.08)		0.67*** (0.08)		0.78*** (0.09)		0.76*** (0.09)
High physical attractiveness			0.30*** (0.07)	0.29*** (0.07)			0.94*** (0.06)	0.93*** (0.06)
Constant	7.09*** (0.08)	6.76*** (0.09)	6.94*** (0.08)	6.61*** (0.09)	6.59*** (0.10)	6.20*** (0.11)	6.13*** (0.11)	5.75*** (0.11)
N (ratings)	6454	6454	6454	6454	6456	6456	6456	6456
N (respondents)	2162	2162	2162	2162	2159	2159	2159	2159

Notes: Results from random intercept multilevel models, effect of religious badges on perceived trustworthiness, exploratory analyses divided by vignette gender. Models 1 and 5 show the main effect for female and male vignettes, respectively; Models 2 and 6 control for social class of the vignette profile, Models 3 and 7 control for facial attractiveness of the vignette profile, Models 4 and 8 control for both social class and facial attractiveness. *$p < .05$. **$p < .01$. ***$p < .001$.

TABLE D.3: Main results of religiosity treatment on perceived trustworthiness

	(1)	(2)	(3)	(4)
Non-religious vignette	ref.	ref.	ref.	ref.
Religious vignette	-0.30***	-0.30***	-0.29***	-0.29***
	(0.06)	(0.06)	(0.06)	(0.06)
Devout vignette	-1.11***	-1.11***	-1.11***	-1.11***
	(0.09)	(0.09)	(0.09)	(0.09)
High social class		0.71***		0.71***
		(0.07)		(0.07)
High physical attractiveness			0.61***	0.60***
			(0.04)	(0.04)
Constant	6.84***	6.49***	6.54***	6.19***
	(0.08)	(0.08)	(0.09)	(0.09)
N (ratings)	12910	12910	12910	12910
N (respondents)	2163	2163	2163	2163

Notes: Results from random intercept multilevel models, effect of religious badges on perceived trustworthiness. Model 1 (pre-registered) shows the main effect, Model 2 controls for social class of the vignette profile, Model 3 controls for facial attractiveness of the vignette profile, Model 4 controls for both social class and facial attractiveness. $*p < .05$. $**p < .01$. $***p < .001$.
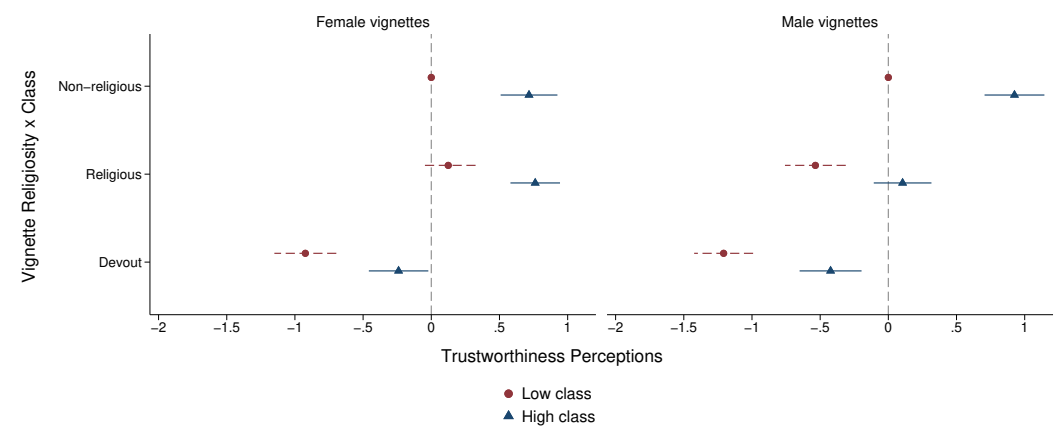
FIGURE D.3: Perceptions of trustworthiness by vignette gender and religiosity treatment (veil or beard), by vignette social class (occupation). $N = 12{,}910$ ratings from 2,162 respondents.

´

TABLE D.4: Perceptions of trustworthiness by vignette religiosity x
social class

	(1) All	(2) All	(3) Female vignettes	(4) Male vignettes
High social class	0.71*** (0.04)			
Non-religious x low class		ref.	ref.	ref.
Non-religious x high class		0.78*** (0.09)	0.71*** (0.11)	0.93*** (0.12)
Religious x low class		-0.25*** (0.07)	0.12 (0.10)	-0.53*** (0.11)
Religious x high class		0.44*** (0.10)	0.77*** (0.10)	0.11 (0.14)
Devout x low class		-1.04*** (0.10)	-0.92*** (0.14)	-1.20*** (0.11)
Devout x high class		-0.39*** (0.12)	-0.25 (0.14)	-0.42** (0.14)
Constant	6.02*** (0.08)	6.45*** (0.09)	6.74*** (0.09)	6.12*** (0.11)
N (ratings)	12910	12910	6454	6456
N (respondents)	2163	2163	2162	2159

Notes: Results from random intercept multilevel models, effect of vignette religiosity and social class on perceived trustworthiness. *$p < .05$. **$p < .01$. ***$p < .001$.

´

TABLE D.5: Perceptions of trustworthiness by vignette facial attractiveness x social class

	(1) All	(2) All	(3) Female vignettes	(4) Male vignettes
High attractiveness	0.60***	0.68***	0.36***	0.99***
	(0.05)	(0.06)	(0.10)	(0.08)
High social class	0.71***	0.79***	0.75***	0.84***
	(0.07)	(0.08)	(0.11)	(0.11)
High attractiveness x high social class		-0.15	-0.15	-0.13
		(0.08)	(0.13)	(0.13)
Constant	5.72***	5.68***	6.30***	5.06***
	(0.09)	(0.10)	(0.11)	(0.12)
N (ratings)	12910	12910	6454	6456
N (respondents)	2163	2163	2162	2159

Notes: Results from random intercept multilevel models, effect of high facial attractiveness and high social class on perceived trustworthiness. *$p < .05$. **$p < .01$. ***$p < .001$.

´

TABLE D.6: Perceptions of trustworthiness by vignette religiosity x
facial attractiveness

	(1) All	(2) Female vignettes	(3) Male vignettes
Non-religious x low attractiveness	ref.	ref.	ref.
Non-religious x high attractiveness	0.75***	0.53***	0.96***
	(0.07)	(0.10)	(0.09)
Religious x low attractiveness	-0.18*	0.22*	-0.64***
	(0.09)	(0.11)	(0.13)
Religious x high attractiveness	0.34***	0.49***	0.24*
	(0.08)	(0.11)	(0.11)
Devout x low attractiveness	-1.01***	-0.74***	-1.30***
	(0.10)	(0.13)	(0.13)
Devout x high attractiveness	-0.46***	-0.60***	-0.32**
	(0.11)	(0.15)	(0.12)
Constant	6.47***	6.83***	6.12***
	(0.09)	(0.09)	(0.12)
N (ratings)	12910	6454	6456
N (respondents)	2163	2162	2159

Notes: Results from random intercept multilevel models, effect of vignette religiosity and social class on perceived trustworthiness. $*p < .05$. $**p < .01$. $***p < .001$.
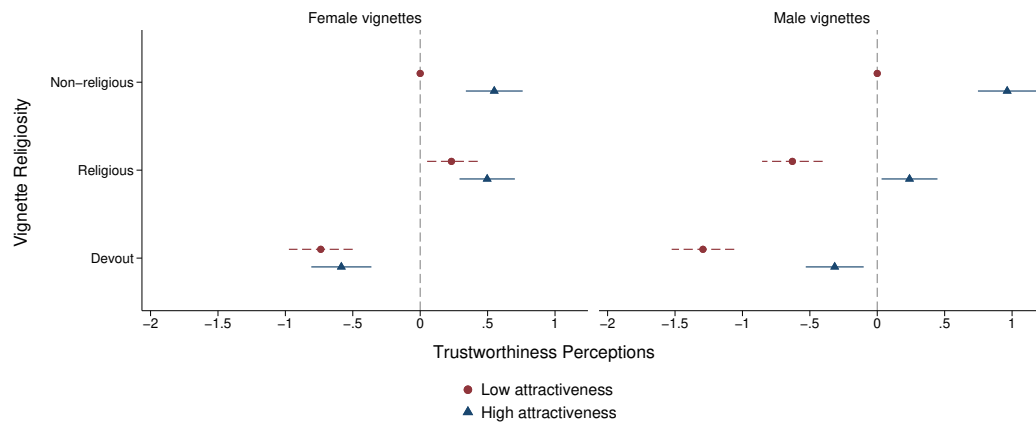
FIGURE D.4: Perceptions of trustworthiness by vignette gender, religiosity treatment (veil or beard), and vignette physical attractiveness. $N = 12{,}910$ ratings from 2,162 respondents.

TABLE D.7: Perceptions of trustworthiness by vignette religiosity and self-reported respondent religiosity.

	(1) All	(2) Female vignettes	(3) Male vignettes
Secular vignette x low religiosity	ref.	ref.	ref.
Secular vignette x high religiosity	0.00	-0.07	0.08
	(0.11)	(0.12)	(0.14)
Religious vignette x low religiosity	-0.57***	-0.24	-0.89***
	(0.11)	(0.13)	(0.14)
Religious vignette x high religiosity	-0.11	0.24*	-0.45**
	(0.11)	(0.12)	(0.15)
Devout vignette x low religiosity	-1.63***	-1.65***	-1.61***
	(0.14)	(0.19)	(0.14)
Devout vignette x high religiosity	-0.77***	-0.54***	-0.99***
	(0.11)	(0.13)	(0.14)
Constant	6.85***	7.14***	6.54***
	(0.10)	(0.10)	(0.12)
N (ratings)	12801	6398	6403
N (respondents)	2141	2140	2141

Notes: Results from random intercept multilevel models, effect of religious badges on perceived trustworthiness. Model 1 pre-registered, Models 2 and 3 exploratory analyses split by vignette gender. *$p < .05$. **$p < .01$. ***$p < .001$.

TABLE D.8: Perceptions of trustworthiness by vignette religiosity and respondents' self-reported wearing of religious badges in public.

	(1) All	(2) Female vignettes	(3) Male vignettes
Non-religious vignette x no badge	ref.	ref.	ref.
Non-religious vignette x wears badge	-0.15 (0.10)	-0.23 (0.12)	-0.05 (0.12)
Religious vignette x no badge	-0.50*** (0.10)	-0.23 (0.12)	-0.77*** (0.14)
Religious vignette x wears badge	-0.31** (0.11)	0.07 (0.13)	-0.68*** (0.14)
Devout vignette x no badge	-1.48*** (0.12)	-1.35*** (0.17)	-1.60*** (0.13)
Devout vignette x wears badge	-1.02*** (0.12)	-0.90*** (0.14)	-1.13*** (0.15)
Constant	6.93*** (0.10)	7.23*** (0.10)	6.63*** (0.12)
N (ratings)	12663	6329	6334
N (respondents)	2119	2118	2118

Notes: Results from random intercept multilevel models. Model 1 includes all vignettes, Models 2 and 3 split by vignette gender. $*p < .05$. $**p < .01$. $***p < .001$.

TABLE D.9: Perceptions of trustworthiness by vignette religiosity and
interviewer rating of respondent religiosity.

	(1) All	(2) Female vignettes	(3) Male vignettes
Non-religious vignette x low religiosity	ref.	ref.	ref.
Non-religious vignette x high religiosity	0.17 (0.10)	0.13 (0.11)	0.22 (0.12)
Religious vignette x low religiosity	-0.45*** (0.09)	-0.10 (0.11)	-0.80*** (0.12)
Religious vignette x high religiosity	0.06 (0.10)	0.44*** (0.12)	-0.32* (0.15)
Devout vignette x low religiosity	-1.40*** (0.12)	-1.33*** (0.16)	-1.48*** (0.12)
Devout vignette x high religiosity	-0.61*** (0.12)	-0.35* (0.14)	-0.87*** (0.15)
Constant	6.77*** (0.09)	7.03*** (0.09)	6.49*** (0.11)
N (ratings)	12765	6381	6384
N (respondents)	2137	2136	2135

Notes: Results from random intercept multilevel models. Model 1 includes all vignettes, Models 2 and 3 split by vignette gender. *$p < .05$. **$p < .01$. ***$p < .001$.

TABLE D.10: Perceptions of trustworthiness by vignette religiosity x self-rated respondent religiosity, including two-way clustering, interviewer fixed effects, and sample weights.

	All		Female vignettes		Male vignettes	
	(1)	(2)	(3)	(4)	(5)	(6)
Secular vignette	ref.		ref.		ref.	
Religious vignette	-0.29***		0.07		-0.65***	
	(0.07)		(0.08)		(0.09)	
Devout vignette	-1.10***		-0.96***		-1.25***	
	(0.09)		(0.12)		(0.10)	
Secular vignette x below mean religiosity		ref.		ref.		ref.
Secular vignette x above mean religiosity		0.04		-0.08		0.16
		(0.11)		(0.13)		(0.13)
Religious vignette x below mean religiosity		-0.63***		-0.31*		-0.95***
		(0.12)		(0.14)		(0.15)
Religious vignette x above mean religiosity		-0.02		0.25		-0.29*
		(0.11)		(0.13)		(0.13)
Devout vignette x below mean religiosity		-1.66***		-1.70***		-1.61***
		(0.15)		(0.20)		(0.14)
Devout vignette x above mean religiosity		-0.70***		-0.55***		-0.85***
		(0.12)		(0.14)		(0.14)
Constant	6.82***	6.80***	7.08***	7.13***	6.57***	6.47***
	(0.05)	(0.08)	(0.06)	(0.10)	(0.06)	(0.10)
N (ratings)	12910	12801	6454	6398	6456	6403
N (respondents)	2163	2141	2162	2140	2159	2141

Notes on Table D.10: Results from high-dimensional fixed effects models including two-way clustering (respondents and neighborhoods), fixed effects for interviewers, and sample weights. $*p < .05$. $**p < .01$. $***p < .001$.
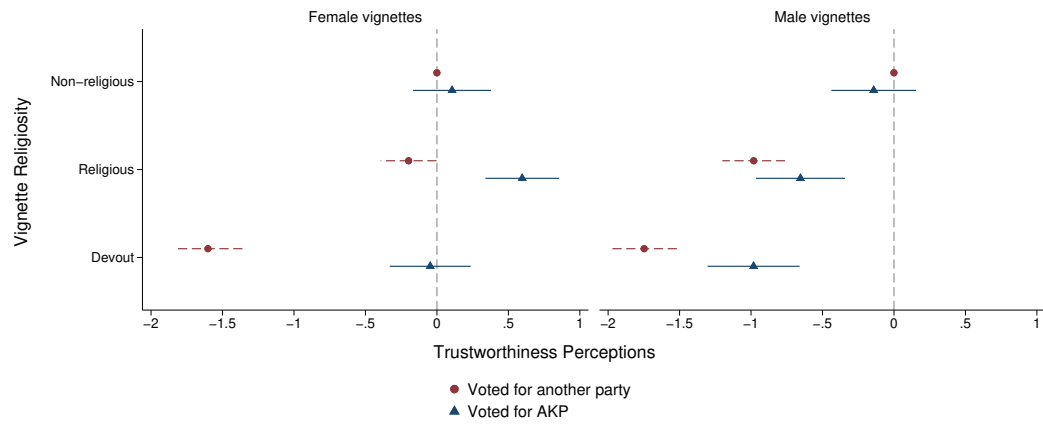

FIGURE D.5: Perceptions of trustworthiness by vignette gender and religiosity treatment (veil or beard), by respondent support of the AKP in the last federal election. $N = 8{,}972$ ratings from 1,500 respondents who voted in the last election.

TABLE D.11: Perceptions of trustworthiness by vignette religiosity x respondents' political party support.

	(1) All	(2) Female vignettes	(3) Male vignettes
Non-religious vignette x other party	ref.	ref.	ref.
Non-religious vignette x AKP voter	0.05 (0.13)	0.06 (0.17)	0.04 (0.14)
Religious vignette x other party	-0.59*** (0.10)	-0.20 (0.12)	-0.98*** (0.12)
Religious vignette x AKP voter	0.04 (0.11)	0.55*** (0.15)	-0.48*** (0.14)
Devout vignette x other party	-1.68*** (0.13)	-1.60*** (0.18)	-1.75*** (0.13)
Devout vignette x AKP voter	-0.45*** (0.13)	-0.09 (0.15)	-0.80*** (0.16)
Constant	6.90*** (0.10)	7.13*** (0.10)	6.67*** (0.12)
N (ratings)	8972	4486	4486
N (respondents)	1500	1500	1499

Notes: Results from random intercept multilevel models, effect of vignette religiosity and respondent political party support on perceived trustworthiness. Model 1 pre-registered, Models 2 and 3 exploratory analyses split by vignette gender. $*p < .05$. $**p < .01$. $***p < .001$.
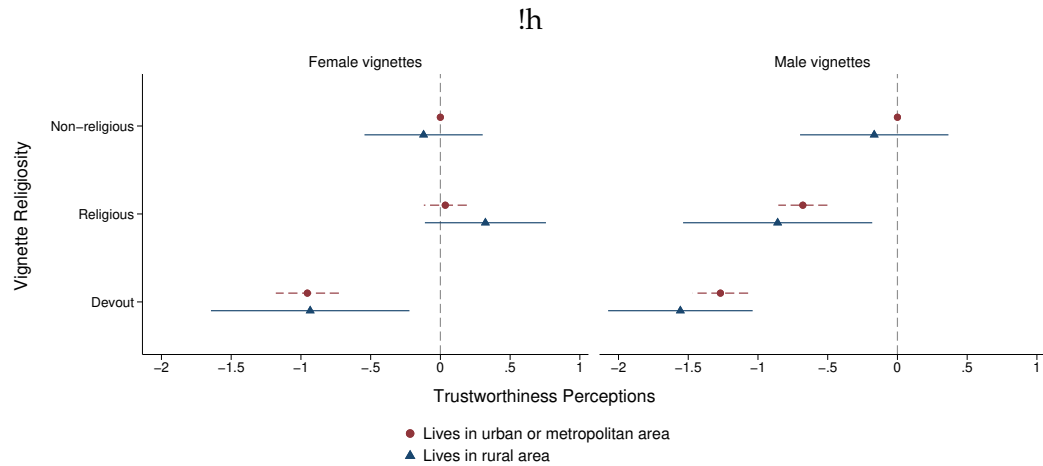
!h

FIGURE D.6: Perceptions of trustworthiness by vignette gender and religiosity treatment (veil or beard), by respondent community size (rural or urban/metropolitan). $N = 12{,}910$ ratings from 2,162 respondents.
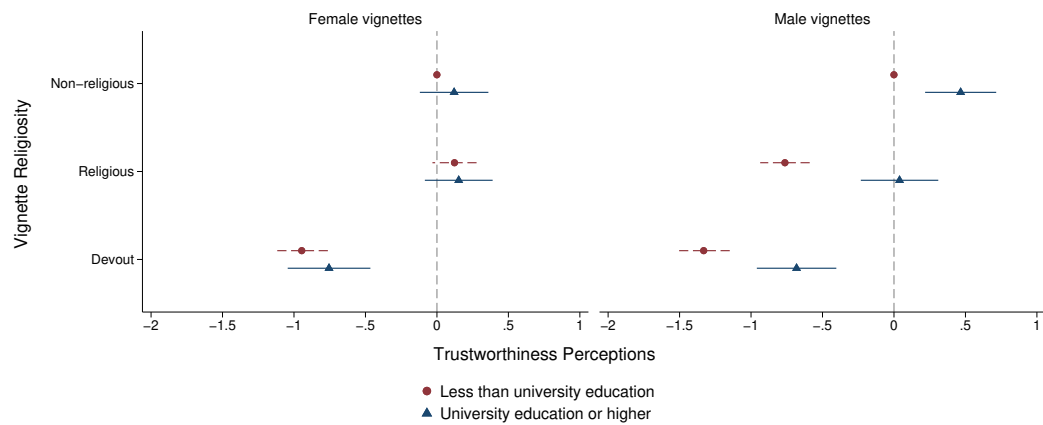

FIGURE D.7: Perceptions of trustworthiness by vignette gender and religiosity treatment (veil or beard), by respondent educational attainment (university degree). $N = 12{,}755$ ratings from 2,133 respondents.

TABLE D.12: Perceptions of trustworthiness by vignette religiosity treatment and controlling for perceived vignette religiosity and political party support.

	Female vignettes					Male vignettes				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Non-religious vignette	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Religious vignette	0.09 (0.08)	-0.67*** (0.08)	0.23** (0.09)	0.25*** (0.05)	-0.19* (0.08)	-0.68*** (0.09)	-0.86*** (0.09)	-0.55*** (0.09)	0.05 (0.07)	-0.03 (0.07)
Devout vignette	-0.94*** (0.11)	-1.88*** (0.11)	-0.75*** (0.12)	-0.38*** (0.07)	-0.97*** (0.11)	-1.28*** (0.10)	-1.56*** (0.10)	-1.10*** (0.09)	0.02 (0.07)	-0.11 (0.07)
Perceived religiosity		0.40*** (0.03)			0.28*** (0.02)		0.34*** (0.03)			0.21*** (0.02)
Perceived support of opposition (CHP) coalition			0.05* (0.02)		0.03* (0.02)			0.09*** (0.02)		0.07*** (0.01)
Perceived attractiveness				0.57*** (0.02)	0.50*** (0.02)				0.66*** (0.02)	0.61*** (0.02)
Constant	7.09*** (0.08)	5.11*** (0.18)	6.78*** (0.17)	3.37*** (0.14)	2.27*** (0.17)	6.59*** (0.10)	4.87*** (0.17)	6.10*** (0.13)	2.50*** (0.13)	1.41*** (0.15)
N (ratings)	6454	6400	6371	6424	6321	6456	6370	6353	6424	6292
N (respondents)	2162	2153	2143	2159	2138	2159	2143	2133	2157	2127
AIC	30888	29829	30496	28797	27914	31665	30798	31131	28973	28172

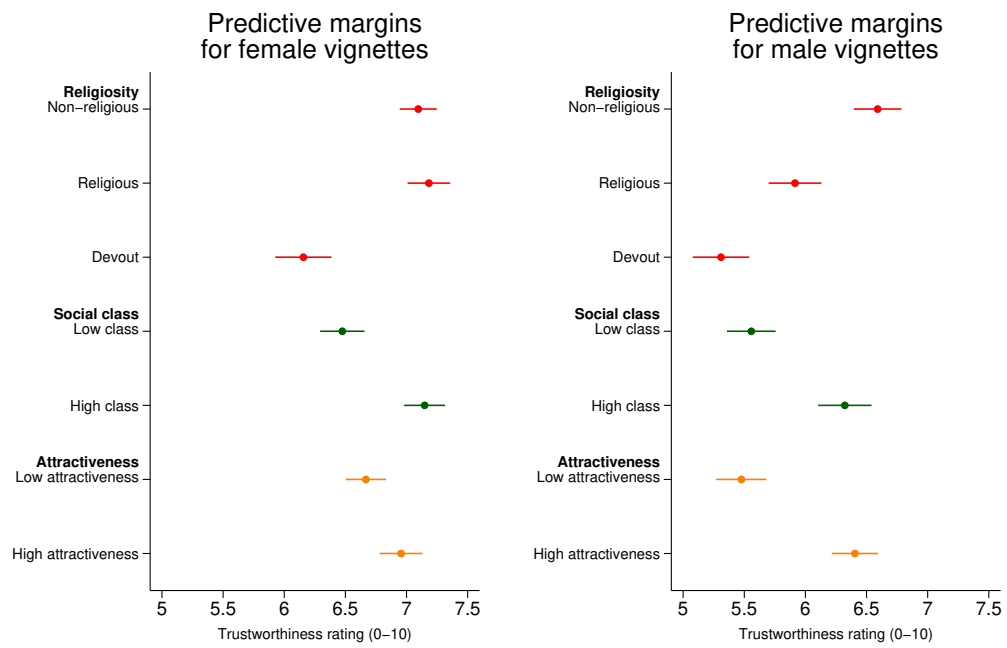*Notes*: Results from random intercept multilevel models. $*p < .05$. $**p < .01$. $***p < .001$.

FIGURE D.8: Perceptions of trustworthiness by gender and vignette dimensions, predictive margins. $N = 12{,}910$ ratings from 2,163 respondents.

Bibliography

Acquisti, A., & Fong, C. (2020). An Experiment in Hiring Discrimination via Online Social Networks. *Management Science*, *66*(3), 1005–1024. https://doi.org/10.1287/mnsc.2018.3269

Adida, C. L., Laitin, D. D., & Valfort, M.-A. (2010). Identifying barriers to Muslim integration in France. *Proceedings of the National Academy of Sciences*, *107*(52), 22384–22390. https://doi.org/10.1073/pnas.1015550107

Adsera, A., & Chiswick, B. R. (2007). Are there gender and country of origin differences in immigrant labor market outcomes across European destinations? *Journal of Population Economics*, *20*(3), 495–526. https://doi.org/10.1007/s00148-006-0082-y

Ahmad, A. (2020). Ethnic discrimination against second-generation immigrants in hiring: Empirical evidence from a correspondence test. *European Societies*, *22*(5), 659–681. https://doi.org/10.1080/14616696.2020.1822536

Ahmed, S., & Gorey, K. M. (2023). Employment discrimination faced by Muslim women wearing the hijab: Exploratory meta-analysis. *Journal of Ethnic & Cultural Diversity in Social Work*, *32*(3), 115–123. https://doi.org/10.1080/15313204.2020.1870601

Aksoy, O., & Gambetta, D. (2016). Behind the veil: The strategic use of religious garb. *European Sociological Review*, *32*(6), 792–806.

Aksoy, O., & Gambetta, D. (2021). The Politics behind the Veil. *European Sociological Review*, *37*(1), 67–88. https://doi.org/10.1093/esr/jcaa035

Aksoy, O., & Wiertz, D. (2024). The impact of religious involvement on trust, volunteering, and perceived cooperativeness: Evidence from two British panels. *European Sociological Review*, *40*(1), 143–159. https://doi.org/10.1093/esr/jcad024

Alba, R. (2005). Bright vs. blurred boundaries: Second-generation assimilation and exclusion in France, Germany, and the United States. *Ethnic and Racial Studies*, *28*(1), 20–49. https://doi.org/10.1080/0141987042000280003

Algan, Y., Dustmann, C., Glitz, A., & Manning, A. (2010). The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom*. *The Economic Journal*, *120*(542), F4–F30. https://doi.org/10.1111/j.1468-0297.2009.02338.x

Alicke, M. D., Smith, R. H., & Klotz, M. L. (1986). Judgments of Physical Attractiveness: The Role of Faces and Bodies. *Personality and Social Psychology Bulletin*, *12*(4), 381–389. https://doi.org/10.1177/0146167286124001

Allmendinger, J. (1989). Educational Systems and Labor Market Outcomes. *European Sociological Review*, *5*(3), 231–250.

Antonczyk, D., Fitzenberger, B., & Sommerfeld, K. (2010). Rising wage inequality, the decline of collective bargaining, and the gender wage gap. *Labour Economics*, *17*(5), 835–847. https://doi.org/10.1016/j.labeco.2010.04.008

Arrow, K. J. (1973). The Theory of Discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in Labor Markets* (pp. 3–33). Princeton University Press. https://doi.org/10.1515/9781400867066

Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology*, *41*(2), 76–81. https://doi.org/10.1027/1864-9335/a000011

Atac, I. E., & Adler, G. J., Jr. (2024). Religious rebound, political backlash, and the youngest cohort: Understanding religious change in Turkey. *Social Forces*, soae102. https://doi.org/10.1093/sf/soae102

Auer, D. (2022). Firing discrimination: Selective labor market responses of firms during the COVID-19 economic crisis. *PLOS ONE*, *17*(1), e0262337. https://doi.org/10.1371/journal.pone.0262337

Auspurg, K., Hinz, T., & Schmid, L. (2017). Contexts and conditions of ethnic discrimination: Evidence from a field experiment in a German housing market. *Journal of Housing Economics*, *35*, 26–36. https://doi.org/10.1016/j.jhe.2017.01.003

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

Aytaç, S. E., & Çarkoğlu, A. (2019). Ethnicity and religiosity-based prejudice in Turkey: Evidence from a survey experiment. *International Political Science Review*, *40*(1), 58–72. https://doi.org/10.1177/0192512117696333

Baert, S. (2018). Facebook profile picture appearance affects recruiters' first hiring decisions. *New Media & Society*, *20*(3), 1220–1239. https://doi.org/10.1177/1461444816687294

Baert, S., & De Pauw, A.-S. (2014). Is ethnic discrimination due to distaste or statistics? *Economics Letters*, *125*(2), 270–273. https://doi.org/10.1016/j.econlet.2014.09.020

Baltes, B. B., & Rudolph, C. W. (2010). Examining the effect of negative Turkish stereotypes on evaluative workplace outcomes in Germany. *Journal of Managerial Psychology*, *25*(2), 148–158. https://doi.org/10.1108/02683941011019357

Basilio, L., Bauer, T. K., & Kramer, A. (2017). Transferability of Human Capital and Immigrant Assimilation: An Analysis for Germany. *LABOUR*, *31*(3), 245–264. https://doi.org/10.1111/labr.12096

Bauer, P. C., & Freitag, M. (2018). Measuring trust. *The Oxford handbook of social and political trust*, *15*.

Bauldry, S., Shanahan, M. J., Russo, R., Roberts, B. W., & Damian, R. (2016). Attractiveness Compensates for Low Status Background in the Prediction of Educational Attainment. *PLOS ONE*, *11*(6), e0155313. https://doi.org/10.1371/journal.pone.0155313

Beaver, K. M., Boccio, C., Smith, S., & Ferguson, C. J. (2019). Physical attractiveness and criminal justice processing: Results from a longitudinal sample of youth and young adults. *Psychiatry, Psychology and Law*, *26*(4), 669–681. https://doi.org/10.1080/13218719.2019.1618750

Becker, G. S. (1971, August). *The Economics of Discrimination*. University of Chicago Press.

Benard, S., Manago, B., Acosta Russian, A., & Cha, Y. (2023). Mapping the Content of Asian Stereotypes in the United States: Intersections with Ethnicity, Gender, Income, and Birthplace. *Social Psychology Quarterly*, *86*(4), 432–456. https://doi.org/10.1177/01902725221126188

Berger, J., Cohen, B. P., & Zelditch, M. (1972). Status Characteristics and Social Interaction. *American Sociological Review*, *37*(3), 241–255. https://doi.org/10.2307/2093465

Berger, J., & Fişek, M. H. (2006). Diffuse Status Characteristics and the Spread of Status Value: A Formal Theory. *American Journal of Sociology*, *111*(4), 1038–1079. https://doi.org/10.1086/498633

Berger, J., Norman, R. Z., Balkwell, J. W., & Smith, R. F. (1992). Status Inconsistency in Task Situations: A Test of Four Status Processing Principles. *American Sociological Review*, *57*(6), 843–855. https://doi.org/10.2307/2096127

Berger, J., Rosenholtz, S. J., & Zelditch, M. (1980). Status Organizing Processes. *Annual Review of Sociology*, *6*, 479–508.

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting Up an Online Panel Representative of the General Population: The German Internet Panel. *Field Methods*, *27*(4), 391–408. https://doi.org/10.1177/1525822X15574494

Boldry, J. G., Gaertner, L., & Quinn, J. (2007). Measuring the Measures: A Meta-Analytic Investigation of the Measures of Outgroup Homogeneity. *Group Processes & Intergroup Relations*, *10*(2), 157–178. https://doi.org/10.1177/1368430207075153

Bonaccolto-Töpfer, M., Castagnetti, C., & Rosti, L. (2023). Changes in the gender pay gap over time: The case of West Germany. *Journal for Labour Market Research*, *57*(1), 11. https://doi.org/10.1186/s12651-023-00338-7

Boon-Falleur, M., André, J.-B., Baumard, N., & Nettle, D. (2024). Household Wealth is Associated With Perceived Trustworthiness in a Diverse Set of Countries. *Social Psychological and Personality Science*, 19485506241289461. https://doi.org/10.1177/19485506241289461

Bouckaert, J., & Dhaene, G. (2004). Inter-ethnic trust and reciprocity: Results of an experiment with small businessmen. *European Journal of Political Economy*, *20*(4), 869–886. https://doi.org/10.1016/j.ejpoleco.2003.08.006

Bourdieu, P. (1984). *Distinction: A Social Critique Of The Judgement Of Taste* (R. Nice, Trans.). Harvard University Press.

Bourdieu, P. (1986). The Forms of Capital. In J. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood.

Bozoyan, C., & Wolbring, T. (2018). The Weight Wage Penalty: A Mechanism Approach to Discrimination. *European Sociological Review, 34*(3), 254–267. https://doi.org/10.1093/esr/jcy009

Browne, I., & Misra, J. (2003). The Intersection of Gender and Race in the Labor Market. *Annual Review of Sociology, 29*(1), 487–513. https://doi.org/10.1146/annurev.soc.29.010202.100016

Brüderl, J., Drobnič, S., Hank, K., Neyer, F. J., Walper, S., Alt, P., Borschel, E., Bozoyan, C., Garrett, M., Geissler, S., Gonzalez Avilés, T., Gröpler, N., Hajek, K., Herzig, M., Huyer-May, B., Lenke, R., Lorenz, R., Lutz, K., Minkus, L., . . . Wetzel, M. (2021). *The German Family Panel (pairfam)* (tech. rep. No. ZA5678 Data file Version 12.0.0). GESIS Data Archive. Cologne. https://doi.org/10.4232/pairfam.5678.12.0.0

Brustkern, J., Heinrichs, M., Walker, M., & Schiller, B. (2021). Facial threat affects trust more strongly than facial attractiveness in women than it does in men. *Scientific Reports, 11*(1), 22475. https://doi.org/10.1038/s41598-021-01775-5

Budde, R., & Eilers, L. (2014). *Sozioökonomische Daten auf Rasterebene: Datenbeschreibung der microm-Rasterdaten* (Research Report No. 77). RWI Materialien. https://www.econstor.eu/handle/10419/97627

Bursell, M. (2014). The Multiple Burdens of Foreign-Named Men—Evidence from a Field Experiment on Gendered Ethnic Hiring Discrimination in Sweden. *European Sociological Review, 30*(3), 399–409. https://doi.org/10.1093/esr/jcu047

Callander, D., Newman, C. E., & Holt, M. (2015). Is Sexual Racism Really Racism? Distinguishing Attitudes Toward Sexual Racism and Generic Racism Among Gay and Bisexual Men. *Archives of Sexual Behavior, 44*(7), 1991–2000. https://doi.org/10.1007/s10508-015-0487-3

Carbon, C.-C., Faerber, S. J., Augustin, M. D., Mitterer, B., & Hutzler, F. (2018). First gender, then attractiveness: Indications of gender-specific attractiveness processing via ERP onsets. *Neuroscience Letters, 686*, 186–192. https://doi.org/10.1016/j.neulet.2018.09.009

Carlsson, M., Fumarco, L., & Rooth, D.-O. (2018). Ethnic discrimination in hiring, labour market tightness and the business cycle - evidence from field experiments. *Applied Economics, 50*(24), 2652–2663. https://doi.org/10.1080/00036846.2017.1406653

Carmines, E., & Nassar, R. (2021). How Social Desirability Bias Affects Immigration Attitudes in a Hyperpolarized Political Environment. *Social Science Quarterly, 102*(4), 1803–1811. https://doi.org/10.1111/ssqu.12982

Carol, S. (2018). Like Will to Like? Partner Choice among Muslim Migrants and Natives in Western Europe. In *Muslims in Europe*. Routledge.

Carvalho, J.-P. (2013). Veiling. *The Quarterly Journal of Economics*, *128*(1), 337–370. https://doi.org/10.1093/qje/qjs045

Cettolin, E., & Suetens, S. (2019). Return on Trust is Lower for Immigrants. *The Economic Journal*, *129*(621), 1992–2009. https://doi.org/10.1111/ecoj.12629

Chan, H. F., Ulrich, F., Altman, H., Schmidt, S. L., Schreyer, D., & Torgler, B. (2022). Beyond performance? The importance of subjective and objective physical appearance in award nominations and receptions in football. *Journal of Economic Behavior & Organization*, *204*, 271–289. https://doi.org/10.1016/j.jebo.2022.10.022

Chiswick, B. R., & Miller, P. W. (2009). The international transferability of immigrants' human capital. *Economics of Education Review*, *28*(2), 162–169. https://doi.org/10.1016/j.econedurev.2008.07.002

Choi, D. D., Poertner, M., & Sambanis, N. (2019). Parochialism, social norms, and discrimination against immigrants. *Proceedings of the National Academy of Sciences*, *116*(33), 16274–16279. https://doi.org/10.1073/pnas.1820146116

Chuah, S. H., Gächter, S., Hoffmann, R., & Tan, J. H. W. (2016). Religion, discrimination and trust across three cultures. *European Economic Review*, *90*, 280–301. https://doi.org/10.1016/j.euroecorev.2016.03.008

Chvaja, R., Chinchilla, J., Goméz, Á., & Lang, M. (2023). Religious costly signal induces more trustworthiness than secular costly signal: A study of pilgrimage to Santiago de Compostela. *European Journal of Social Psychology*, *53*(6), 1294–1308. https://doi.org/10.1002/ejsp.2975

Çokgezen, M. (2022). Can the State Make you More Religious? Evidence from Turkish Experience. *Journal for the Scientific Study of Religion*, *61*(2), 349–373. https://doi.org/10.1111/jssr.12785

Coleman, J. S. (1986). Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology*, *91*(6), 1309–1335.

Combes, P.-P., Decreuse, B., Laouénan, M., & Trannoy, A. (2016). Customer Discrimination and Employment Outcomes: Theory and Evidence from the French Labor Market. *Journal of Labor Economics*, *34*(1), 107–160. https://doi.org/10.1086/682332

Cook, K. (2001). *Trust in Society*. Russell Sage Foundation.

Correia, S. (2016). *Linear models with high-dimensional fixed effects: An efficient and feasible estimator* (tech. rep.).

Correia, S., Guimarães, P., & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, *20*(1), 95–115. https://doi.org/10.1177/1536867X20909691

Correll, S. J., & Ridgeway, C. L. (2006). Expectation States Theory. In J. Delamater (Ed.), *Handbook of Social Psychology* (pp. 29–51). Springer US. https://doi.org/10.1007/0-387-36921-X_2

Costa, P. L., McDuffie, J. W., Brown, S. E. V., He, Y., Ikner, B. N., Sabat, I. E., & Miner, K. N. (2023). Microaggressions: Mega problems or micro issues? A meta-analysis. *Journal of Community Psychology*, *51*(1), 137–153. https://doi.org/10.1002/jcop.22885

Cox, J. C., & Orman, W. H. (2015). Trust and trustworthiness of immigrants and native-born Americans. *Journal of Behavioral and Experimental Economics*, *57*, 1–8. https://doi.org/10.1016/j.socec.2015.03.008

Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, *1989*(1), 139–168. https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8/

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. *Advances in Experimental Social Psychology*, *40*, 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0

Dahl, M., & Krog, N. (2018). Experimental Evidence of Discrimination in the Labour Market: Intersections between Ethnicity, Gender, and Socio-Economic Status. *European Sociological Review*, *34*(4), 402–417. https://doi.org/10.1093/esr/jcy020

Darity, W. A. (2003). Employment Discrimination, Segregation, and Health. *American Journal of Public Health*, *93*(2), 226–231. https://doi.org/10.2105/AJPH.93.2.226

De Neys, W., Hopfensitz, A., & Bonnefon, J.-F. (2017). Split-Second Trustworthiness Detection From Faces in an Economic Game. *Experimental Psychology*, *64*(4), 231–239. https://doi.org/10.1027/1618-3169/a000367

DeLamater, J. D., & Hyde, J. S. (1998). Essentialism vs. social constructionism in the study of human sexuality. *The Journal of Sex Research*, *35*(1), 10–18. https://doi.org/10.1080/00224499809551913

Deryugina, T., & Shurchkov, O. (2015). Now you see it, now you don't: The vanishing beauty premium. *Journal of Economic Behavior & Organization*, *116*, 331–345. https://doi.org/10.1016/j.jebo.2015.05.007

Di Stasio, V., Lancee, B., Veit, S., & Yemane, R. (2021). Muslim by default or religious discrimination? Results from a cross-national field experiment on hiring discrimination. *Journal of Ethnic and Migration Studies*, *47*(6), 1305–1326. https://doi.org/10.1080/1369183X.2019.1622826

Di Stasio, V., & Larsen, E. N. (2020). The Racialized and Gendered Workplace: Applying an Intersectional Lens to a Field Experiment on Hiring Discrimination in Five European Labor Markets. *Social Psychology Quarterly*, *83*(3), 229–250. https://doi.org/10.1177/0190272520902994

Dias, F. A. (2020). How skin color, class status, and gender intersect in the labor market: Evidence from a field experiment. *Research in Social Stratification and Mobility*, *65*, 100477. https://doi.org/10.1016/j.rssm.2020.100477

Dietrich, B. J., & Sands, M. L. (2023). Seeing racial avoidance on New York City streets. *Nature Human Behaviour*, *7*(8), 1275–1281. https://doi.org/10.1038/s41562-023-01589-7

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290. https://doi.org/10.1037/h0033731

Dixon, A. R., & Telles, E. E. (2017). Skin Color and Colorism: Global Research, Concepts, and Measurement. *Annual Review of Sociology*, *43*, 405–424. https://doi.org/10.1146/annurev-soc-060116-053315

Dixson, B. J., & Vasey, P. L. (2012). Beards augment perceptions of men's age, social status, and aggressiveness, but not attractiveness. *Behavioral Ecology*, *23*(3), 481–490. https://doi.org/10.1093/beheco/arr214

Dixson, B. J. W., Barkhuizen, C. L., & Craig, B. M. (2021). Beards Increase the Speed, Accuracy, and Explicit Judgments of Facial Threat. *Adaptive Human Behavior and Physiology*, *7*(4), 347–362. https://doi.org/10.1007/s40750-021-00169-1

Dong, Y., Li, Y., & Sun, T. (2014). Happy faces considered trustworthy irrespective of perceiver's mood: Challenges to the mood congruency effect. *Computers & Security*, *47*, 85–93. https://doi.org/10.1016/j.cose.2014.04.010

Doorley, K., & Sierminska, E. (2015). Myth or fact? The beauty premium across the wage distribution in Germany. *Economics Letters*, *129*, 29–34. https://doi.org/10.1016/j.econlet.2015.01.033

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Eckel, C. C., & Petrie, R. (2011). Face Value. *American Economic Review*, *101*(4), 1497–1513. https://doi.org/10.1257/aer.101.4.1497

Fernández-Reino, M., Di Stasio, V., & Veit, S. (2023). Discrimination unveiled: A field experiment on the barriers faced by Muslim women in Germany, the Netherlands, and Spain. *European Sociological Review*, *39*(3), 479–497. https://doi.org/10.1093/esr/jcac032

Fiala, V., Třebický, V., Pazhoohi, F., Leongómez, J. D., Tureček, P., Saribay, S. A., Akoko, R. M., & Kleisner, K. (2021). Facial attractiveness and preference of sexual dimorphism: A comparison across five populations. *Evolutionary Human Sciences*, *3*, e38. https://doi.org/10.1017/ehs.2021.33

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fletcher, J. M. (2013). The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior & Organization*, *89*, 122–135. https://doi.org/10.1016/j.jebo.2013.02.004

Forster, A. G., & Neugebauer, M. (2024). Factorial Survey Experiments to Predict Real-World Behavior: A Cautionary Tale from Hiring Studies. *Sociological Science*, *11*, 886–906. https://doi.org/10.15195/v11.a32

Fourgassie, L., Subra, B., & Sanitioso, R. B. (2024). Stereotype Content of North African Men and Women in France and Its Relation to Aggression. *Social Psychology Quarterly*, *87*(1), 64–83. https://doi.org/10.1177/01902725231159938

Frederick, D., Forbes, M., Jenkins, B., Reynolds, T., & Walters, T. (2015). Beauty standards. In P. Whelehan & A. Bolin (Eds.), *The international encyclopedia of human sexuality* (pp. 113–196). John Wiley & Sons. https://onlinelibrary.wiley.com/doi/10.1002/9781118896877.wbiehs046

French, M. T. (2002). Physical appearance and earnings: Further evidence. *Applied Economics*, *34*(5), 569–572. https://doi.org/10.1080/00036840010027568

Friberg, J. H., & Midtbøen, A. H. (2018). Ethnicity as skill: Immigrant employment hierarchies in Norwegian low-wage labour markets. *Journal of Ethnic and Migration Studies*, *44*(9), 1463–1478. https://doi.org/10.1080/1369183X.2017.1388160

Friedberg, R. M. (2000). You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics*, *18*(2), 221–251. https://doi.org/10.1086/209957

Froehlich, L., & Schulte, I. (2019). Warmth and competence stereotypes about immigrant groups in Germany. *PLOS ONE*, *14*(9), e0223103. https://doi.org/10.1371/journal.pone.0223103

Gaddis, S. M. (2018). An Introduction to Audit Studies in the Social Sciences. In S. M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 3–44). Springer International Publishing. https://doi.org/10.1007/978-3-319-71153-9_1

Galarza, F. B., & Yamada, G. (2014). Labor Market Discrimination in Lima, Peru: Evidence from a Field Experiment. *World Development*, *58*, 83–94. https://doi.org/10.1016/j.worlddev.2014.01.003

Gambetta, D. (2011). Signaling. In P. Bearman & P. Hedström (Eds.), *The Oxford Handbook of Analytical Sociology* (pp. 168–194). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199215362.013.8

Gambetta, D., & Morisi, D. (2022). COVID-19 infection induces higher trust in strangers. *Proceedings of the National Academy of Sciences*, *119*(32), e2116818119. https://doi.org/10.1073/pnas.2116818119

Garza, R., Afhami, R., Mora, J., & Pazhoohi, F. (2023). Perceptions of Beardedness for Attractiveness, Masculinity, Fighting Ability, and Partner Quality: A cross-cultural Examination Among Hispanic and Iranian Women. *Adaptive Human Behavior and Physiology*, *9*(4), 325–343. https://doi.org/10.1007/s40750-023-00225-y

Gehrsitz, M. (2014). Looks and Labor: Do Attractive People Work More? *LABOUR*, *28*(3), 269–287. https://doi.org/10.1111/labr.12035

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gereke, J., Hellyer, J., Behnert, J., Exner, S., Herbel, A., Jäger, F., Lajic, D., Mezenský, Š., Ngoc Anh, V., Ogłaza, T., Schabinger, J., Sokolova, A., Szafran, D., Tirolf, N., Veit, S., & Zhang, N. (2022). Demographic Change and Group Boundaries in Germany: The Effect of Projected Demographic Decline on Perceptions of Who Has a Migration Background. *Sociological Science*, *9*, 206–220. https://doi.org/10.15195/v9.a9

Gereke, J., & Ruedin, D. (2023). Shared Nationality in Social Exchange: A Trust Vignette Experiment in the United States, South Africa, and Switzerland. *Socius*, *9*, 23780231231189945. https://doi.org/10.1177/23780231231189945

Gereke, J., Schaub, M., & Baldassarri, D. (2020). Gendered Discrimination Against Immigrants: Experimental Evidence. *Frontiers in Sociology*, *5*. https://www.frontiersin.org/articles/10.3389/fsoc.2020.00059

German Internet Panel, Universität Mannheim. (2024). German Internet Panel, Wave 70 (March 2024).

Ghavami, N., & Peplau, L. A. (2013). An Intersectional Analysis of Gender and Ethnic Stereotypes: Testing Three Hypotheses. *Psychology of Women Quarterly*, *37*(1), 113–127. https://doi.org/10.1177/0361684312464203

Ghumman, S., & Ryan, A. M. (2013). Not welcome here: Discrimination towards women who wear the Muslim headscarf. *Human Relations*, *66*(5), 671–698. https://doi.org/10.1177/0018726712469540

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Prentice-Hall.

Gordon, R. A., Crosnoe, R., & Wang, X. (2013). Physical Attractiveness and the Accumulation of Social and Human Capital in Adolescence and Young Adulthood: Assets and Distractions. *Monographs of the Society for Research in Child Development*, *78*(6), 1–137. https://doi.org/10.1002/mono.12060

Goulão, C., Lacomba, J. A., Lagos, F., & Rooth, D.-O. (2024). Weight, attractiveness, and gender when hiring: A field experiment in Spain. *Journal of Economic Behavior & Organization*, *218*, 132–145. https://doi.org/10.1016/j.jebo.2023.11.028

Graham, J., & Haidt, J. (2010). Beyond beliefs: Religions bind individuals into moral communities. *Personality and social psychology review*, *14*(1), 140–150.

Gutiérrez-García, A., Beltrán, D., & Calvo, M. G. (2019). Facial attractiveness impressions precede trustworthiness inferences: Lower detection thresholds and faster decision latencies. *Cognition and Emotion*, *33*(2), 378–385. https://doi.org/10.1080/02699931.2018.1444583

Hakim, C. (2010). Erotic Capital. *European Sociological Review*, *26*(5), 499–518. https://doi.org/10.1093/esr/jcq014

Hall, D. L., Cohen, A. B., Meyer, K. K., Varley, A. H., & Brewer, G. A. (2015). Costly signaling increases trust, even across religious affiliations. *Psychological Science*, *26*(9), 1368–1376.

Hall, P. A., & Soskice, D. (2001). An Introduction to Varieties of Capitalism. In P. A. Hall & D. Soskice (Eds.), *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage* (pp. 1–70). Oxford University Press. https://doi.org/10.1093/0199247757.003.0001

Hamermesh, D. S. (2011). *Beauty Pays: Why Attractive People Are More Successful.* Princeton University Press.

Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the Labor Market. *The American Economic Review*, *84*(5), 1174–1194. Retrieved January 9, 2023, from https://www.jstor.org/stable/2117767

Hansen, K., Rakić, T., & Steffens, M. C. (2018). Foreign-Looking Native-Accented People: More Competent When First Seen Rather Than Heard. *Social Psychological and Personality Science*, *9*(8), 1001–1009. https://doi.org/10.1177/1948550617732389

Hansson, K., Habibnia, H., Goetze, M., & Fiedler, S. (2024). The beauty of prosocial behavior: The bi-directional link between attractiveness and prosocial behavior. *Journal of Economic Behavior & Organization*, *219*, 305–317. https://doi.org/10.1016/j.jebo.2024.01.013

Hardin, R. (2002). *Trust and trustworthiness*. Russell Sage Foundation.

Harper, B. (2000). Beauty, Stature and the Labour Market: A British Cohort Study. *Oxford Bulletin of Economics and Statistics*, *62*(s1), 771–800. https://doi.org/10.1111/1468-0084.0620s1771

Harteveld, E., & Ivarsflaten, E. (2018). Why Women Avoid the Radical Right: Internalized Norms and Party Reputations. *British Journal of Political Science*, *48*(2), 369–384. https://doi.org/10.1017/S0007123415000745

Heath, A. F., Rothon, C., & Kilpi, E. (2008). The Second Generation in Western Europe: Education, Unemployment, and Occupational Attainment. *Annual Review of Sociology*, *34*(1), 211–235. https://doi.org/10.1146/annurev.soc.34.040507.134728

Heilman, M. E., & Saruwatari, L. R. (1979). When beauty is beastly: The effects of appearance and sex on evaluations of job applicants for managerial and non-managerial jobs. *Organizational Behavior and Human Performance*, *23*(3), 360–372. https://doi.org/10.1016/0030-5073(79)90003-5

Heiserman, N. (2023). Intersectional Complexity in Stereotype Content. *American Journal of Sociology*, *129*(3), 753–795. https://doi.org/10.1086/728051

Helbling, M. (2014). Opposing Muslims and the Muslim Headscarf in Western Europe. *European Sociological Review*, *30*(2), 242–257. https://doi.org/10.1093/esr/jct038

Helbling, M., Jäger, F., & Traunmüller, R. (2022). Muslim bias or fear of fundamentalism? A survey experiment in five Western European democracies. *Research & Politics*, *9*(1), 20531680221088491.

Hellyer, J. (2024). Facial attractiveness does not modify the perceived trustworthiness of ethnic minority men. *Scientific Reports*, *14*(1), 27093. https://doi.org/10.1038/s41598-024-78291-9

Hellyer, J., Hellriegel, E., Gereke, J., & Schunck, R. (2023). Pretty unequal? Immigrant-native differences in returns to physical attractiveness in Germany. *Journal of Economic Behavior & Organization*, *215*, 107–119. https://doi.org/10.1016/j.jebo.2023.09.012

Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and human behavior*, *30*(4), 244–260.

Hernan, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.

Hernán, M. A., & Cole, S. R. (2009). Invited Commentary: Causal Diagrams and Measurement Bias. *American Journal of Epidemiology*, *170*(8), 959–962. https://doi.org/10.1093/aje/kwp293

Heyne, S., Kuhlemann, J., & Kogan, I. (2025). Partnership Preferences in Modern Migration Societies: Religious Homophily Among Young Muslims and Christians in Germany. *European Journal of Population*, *41*(1), 4. https://doi.org/10.1007/s10680-024-09728-0

Hill, E. M., Nocks, E. S., & Gardner, L. (1987). Physical attractiveness: Manipulation by physique and status displays. *Ethology and Sociobiology*, *8*(2), 143–154. https://doi.org/10.1016/0162-3095(87)90037-9

Hill, M. E. (2002). Skin Color and the Perception of Attractiveness among African Americans: Does Gender Make a Difference? *Social Psychology Quarterly*, *65*(1), 77–91. https://doi.org/10.2307/3090169

Hofmann, A., & Hallsteinsdóttir, E. (2016). Deutsch-dänische Stereotypenwelten im SMiK-Projekt. *Linguistik online*, *79*(5), 323–345. https://www.redalyc.org/articulo.oa?id=664573573011

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The Effects of Physical Attractiveness on Job-Related Outcomes: A Meta-Analysis of Experimental Studies. *Personnel Psychology*, *56*(2), 431–462. https://doi.org/https://doi.org/10.1111/j.1744-6570.2003.tb00157.x

Huinink, J., Brüderl, J., Nauck, B., Walper, S., Castiglioni, L., & Feldhaus, M. (2011). Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design: Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Konzeptioneller Rahmen und Forschungsdesign. *Journal of Family Research*, *23*(1), 77–101. https://doi.org/10.20377/jfr-235

Hutchings, R. J., Freiburger, E., Sim, M., & Hugenberg, K. (2024). Racial Prejudice Affects Representations of Facial Trustworthiness. *Psychological Science*, *35*(3), 263–276. https://doi.org/10.1177/09567976231225094

Ingwersen, K., & Thomsen, S. L. (2021). The immigrant-native wage gap in Germany revisited. *The Journal of Economic Inequality*, *19*(4), 825–854. https://doi.org/10.1007/s10888-021-09493-8

Jäckle, S., & Metz, T. (2017). Beauty Contest Revisited: The Effects of Perceived Attractiveness, Competence, and Likability on the Electoral Success of German MPs. *Politics & Policy*, *45*(4), 495–534. https://doi.org/10.1111/polp.12209

Jackson, J. C., & Gray, K. (2019). When a good god makes bad people: Testing a theory of religion and immorality. *Journal of personality and social psychology*, *117*(6), 1203.

Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review. *Social Psychology Quarterly*, *58*(2), 108–122. https://doi.org/10.2307/2787149

Jæger, M. M. (2011). "A Thing of Beauty is a Joy Forever"? Returns to Physical Attractiveness over the Life Course. *Social Forces*, *89*(3), 983–1003. https://doi.org/10.1093/sf/89.3.983

Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019). Explaining the persistent influence of facial cues in social decision-making. *Journal of Experimental Psychology: General*, *148*(6), 1008–1021. https://doi.org/10.1037/xge0000591

Jaeger, B., Oud, B., Williams, T., Krumhuber, E. G., Fehr, E., & Engelmann, J. B. (2022). Can people detect the trustworthiness of strangers based on their facial appearance? *Evolution and Human Behavior*, *43*(4), 296–303. https://doi.org/10.1016/j.evolhumbehav.2022.04.004

Jasinskaja-Lahti, I., Liebkind, K., & Perhoniemi, R. (2007). Perceived ethnic discrimination at work and well-being of immigrants in Finland: The moderating role of employment status and work-specific group-level control beliefs. *International Journal of Intercultural Relations*, *31*(2), 223–242. https://doi.org/10.1016/j.ijintrel.2006.02.003

Johnson, B. D., & King, R. D. (2017). Facial Profiling: Race, Physical Appearance, and Punishment. *Criminology*, *55*(3), 520–547. https://doi.org/10.1111/1745-9125.12143

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*(1), 116–131. https://doi.org/10.1037/a0025335

Johnson, S. K., Podratz, K. E., Dipboye, R. L., & Gibbons, E. (2010). Physical Attractiveness Biases in Ratings of Employment Suitability: Tracking Down the "Beauty is Beastly" Effect. *The Journal of Social Psychology*, *150*(3), 301–318. https://doi.org/10.1080/00224540903365414

Jordan, T. R., Yekani, H. A. K., & Sheen, M. (2020). Further investigation of the effects of wearing the hijab: Perception of female facial attractiveness by Emirati Muslim men living in their native Muslim country. *PLOS ONE*, *15*(10), e0239419. https://doi.org/10.1371/journal.pone.0239419

Judge, T. A., Hurst, C., & Simon, L. S. (2009). Does it pay to be smart, attractive, or confident (or all three)? Relationships among general mental ability, physical attractiveness, core self-evaluations, and income. *Journal of Applied Psychology*, *94*(3), 742–755. https://doi.org/10.1037/a0015497

Kaas, L., & Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, *13*(1), 1–20. https://doi.org/10.1111/j.1468-0475.2011.00538.x

Kanazawa, S. (2011). Intelligence and physical attractiveness. *Intelligence*, *39*(1), 7–14. https://doi.org/10.1016/j.intell.2010.11.003

Kang, S. K., & Bodenhausen, G. V. (2015). Multiple Identities in Social Perception and Interaction: Challenges and Opportunities. *Annual Review of Psychology*, *66*, 547–574. https://doi.org/10.1146/annurev-psych-010814-015025

Kanitsar, G. (2023). Leaving the bike unlocked: Trust discrimination in inter-ethnic encounters. *European Sociological Review*, jcad045. https://doi.org/10.1093/esr/jcad045

Kelly, J. M., Kramer, S. R., & Shariff, A. F. (2024). Religiosity predicts prosociality, especially when measured by self-report: A meta-analysis of almost 60 years of research. *Psychological Bulletin*. https://doi.org/10.1037/bul0000413

Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. H. J. (2016). Do we spontaneously form stable trustworthiness impressions from facial appearance? *Journal of Personality and Social Psychology*, *111*(5), 655–664. https://doi.org/10.1037/pspa0000062

Klebl, C., Rhee, J. J., Greenaway, K. H., Luo, Y., & Bastian, B. (2022). Beauty Goes Down to the Core: Attractiveness Biases Moral Character Attributions. *Journal of Nonverbal Behavior*, *46*(1), 83–97. https://doi.org/10.1007/s10919-021-00388-w

Kočnar, T., Saribay, S. A., & Kleisner, K. (2019). Perceived attractiveness of Czech faces across 10 cultures: Associations with sexual shape dimorphism, averageness, fluctuating asymmetry, and eye color. *PLOS ONE*, *14*(11), e0225549. https://doi.org/10.1371/journal.pone.0225549

Kogan, I. (2004). Last Hired, First Fired? The Unemployment Dynamics of Male Immigrants in Germany. *European Sociological Review*, *20*(5), 445–461. https://doi.org/10.1093/esr/jch037

Kogan, I. (2007). A study of immigrants' employment careers in West Germany using the sequence analysis technique. *Social Science Research*, *36*(2), 491–511. https://doi.org/10.1016/j.ssresearch.2006.03.004

Koopmans, R., Veit, S., & Yemane, R. (2019). Taste or statistics? A correspondence study of ethnic, racial and religious labour market discrimination in Germany. *Ethnic and Racial Studies*, *42*(16), 233–252. https://doi.org/10.1080/01419870.2019.1654114

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly*, *72*(5), 847–865. https://doi.org/10.1093/poq/nfn063

Kristen, C., & Granato, N. (2007). The educational attainment of the second generation in Germany: Social origins and ethnic inequality. *Ethnicities*, *7*(3), 343–366. https://doi.org/10.1177/1468796807080233

Kühn, J., & Wolbring, T. (2024). Beauty pays, but not under all circumstances: Evidence on gendered hiring discrimination from a novel experimental treatment using deepfakes. *Research in Social Stratification and Mobility*, *94*, 100992. https://doi.org/10.1016/j.rssm.2024.100992

Kukkonen, I., Pajunen, T., Sarpila, O., & Åberg, E. (2024). Is beauty-based inequality gendered? A systematic review of gender differences in socioeconomic outcomes of physical attractiveness in labor markets. *European Societies*, *26*(1), 117–148. https://doi.org/10.1080/14616696.2023.2210202

Kunst, J. R., Kirkøen, J., & Mohamdain, O. (2023). Hacking attractiveness biases in hiring? The role of beautifying photo-filters. *Management Decision*, *61*(4), 924–943. https://doi.org/10.1108/MD-06-2021-0747

Lacroix, J., Ruedin, D., & Zschirnt, E. (2023). Discrimination driven by variation in social and economic conservatism: Evidence from a nationwide field experiment. *European Sociological Review*, *39*(3), 464–478. https://doi.org/10.1093/esr/jcac051

Lancee, B., Birkelund, G., Coenders, M., Di Stasio, V., Fernandez Reino, M., Heath, A., Koopmans, R., Larsen, E., Polavieja, J. G., Ramos, M., Thijssen, L., Veit, S., Yemane, R., & Zwier, D. (2019). The GEMM Study: A Cross-National Harmonized Field Experiment on Labour Market Discrimination: Technical Report. https://doi.org/10.2139/ssrn.3398191

Landesvatter, C., & Bauer, P. C. (2024). How Valid Are Trust Survey Measures? New Insights From Open-Ended Probing Data and Supervised Machine Learning. *Sociological Methods & Research*, 00491241241234871. https://doi.org/10.1177/00491241241234871

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390–423. https://doi.org/10.1037/0033-2909.126.3.390

Larsen, E. N. (2020). Induced competition in matched correspondence tests: Conceptual and methodological considerations. *Research in Social Stratification and Mobility*, *65*, 100475. https://doi.org/10.1016/j.rssm.2020.100475

Lee, T. L., & Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: Immigrants in the Stereotype Content Model. *International Journal of Intercultural Relations*, *30*(6), 751–768. https://doi.org/10.1016/j.ijintrel.2006.06.005

Leger, K., Dong, J., DeBruine, L. M., Jones, B. C., & Shiramizu, V. K. M. (2023). Assessing the roles of shape prototypicality and sexual dimorphism in ratings

of the trustworthiness of faces. *Scientific Reports*, *13*(1), 15662. https://doi.org/10.1038/s41598-023-42990-6

Lehmer, F., & Ludsteck, J. (2011). The Immigrant Wage Gap in Germany: Are East Europeans Worse Off? *International Migration Review*, *45*(4), 872–906. https://www.jstor.org/stable/41427972

Leicht, K. T. (2008). Broken Down by Race and Gender? Sociological Explanations of New Sources of Earnings Inequality. *Annual Review of Sociology*, *34*(1), 237–255. https://doi.org/10.1146/annurev.soc.34.040507.134627

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The Origins of Social Categorization. *Trends in Cognitive Sciences*, *21*(7), 556–568. https://doi.org/10.1016/j.tics.2017.04.004

Ling, L., Luo, D., & She, G. (2019). Judging a book by its Cover: The influence of physical attractiveness on the promotion of regional leaders. *Journal of Economic Behavior & Organization*, *158*, 1–14. https://doi.org/10.1016/j.jebo.2019.01.005

Lippens, L., Baert, S., Ghekiere, A., Verhaeghe, P.-P., & Derous, E. (2022). Is labour market discrimination against ethnic minorities better explained by taste or statistics? A systematic review of the empirical evidence. *Journal of Ethnic and Migration Studies*, *48*(17), 4243–4276. https://doi.org/10.1080/1369183X.2022.2050191

Lippens, L., Vermeiren, S., & Baert, S. (2023). The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. *European Economic Review*, *151*, 104315. https://doi.org/10.1016/j.euroecorev.2022.104315

Loosveldt, G., & Beullens, K. (2013). 'How long will it take?' An analysis of interview length in the fifth round of the European Social Survey. *Survey Research Methods*, *7*(2), 69–78. https://doi.org/10.18148/srm/2013.v7i2.5086

López Bóo, F., Rossi, M. A., & Urzúa, S. S. (2013). The labor market return to an attractive face: Evidence from a field experiment. *Economics Letters*, *118*(1), 170–172. https://doi.org/10.1016/j.econlet.2012.10.016

López Ortega, A., & Radojevic, M. (2024). Visual Conjoint vs. Text Conjoint and the Differential Discriminatory Effect of (Visible) Social Categories. *Political Behavior*. https://doi.org/10.1007/s11109-024-09953-7

Lundborg, P., Nystedt, P., & Rooth, D.-O. (2014). Height and Earnings: The Role of Cognitive and Noncognitive Skills. *Journal of Human Resources*, *49*(1), 141–166. https://doi.org/10.3368/jhr.49.1.141

Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology*, *47*(2), 391–396. https://doi.org/10.1016/j.jesp.2010.11.002

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Maestripieri, D., Henry, A., & Nickels, N. (2017). Explaining financial and prosocial biases in favor of attractive people: Interdisciplinary perspectives from economics, social psychology, and evolutionary psychology. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X16000340

Marques, J. M., & Paez, D. (1994). The 'Black Sheep Effect': Social Categorization, Rejection of Ingroup Deviates, and Perception of Group Variability. *European Review of Social Psychology*, *5*(1), 37–68. https://doi.org/10.1080/14792779543000011

Marques, J. M., Yzerbyt, V. Y., & Leyens, J.-P. (1988). The "Black Sheep Effect": Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, *18*(1), 1–16. https://doi.org/10.1002/ejsp.2420180102

Marquis, C., Tilcsik, A., & Zhang, Y. (2024). Attractiveness and Attainment: Status, Beauty, and Jobs in China and the United States. *American Journal of Sociology*, *129*(6), 1720–1762. https://doi.org/10.1086/729834

Matsangidou, M., & Otterbacher, J. (2019). What Is Beautiful Continues to Be Good. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human-Computer Interaction – INTERACT 2019* (pp. 243–264). Springer International Publishing. https://doi.org/10.1007/978-3-030-29390-1_14

Maurer-Fazio, M., & Lei, L. (2015). "As rare as a panda": How facial attractiveness, gender, and occupation affect interview callbacks at Chinese firms. *International Journal of Manpower*, *36*(1), 68–85. https://doi.org/10.1108/IJM-12-2014-0258

Mazzella, R., & Feingold, A. (1994). The Effects of Physical Attractiveness, Race, Socioeconomic Status, and Gender of Defendants and Victims on Judgments of Mock Jurors: A Meta-Analysis. *Journal of Applied Social Psychology*, *24*(15), 1315–1338. https://doi.org/10.1111/j.1559-1816.1994.tb01552.x

McCall, L. (2005). The Complexity of Intersectionality. *Signs: Journal of Women in Culture and Society*, *30*(3), 1771–1800. https://doi.org/10.1086/426800

McCullough, M. E., Swartwout, P., Shaver, J. H., Carter, E. C., & Sosis, R. (2016). Christian religious badges instill trust in Christian and non-Christian perceivers. *Psychology of Religion and Spirituality*, *8*(2), 149.

Mears, A. (2014). Aesthetic Labor for the Sociologies of Work, Gender, and Beauty. *Sociology Compass*, *8*(12), 1330–1343. https://doi.org/10.1111/soc4.12211

Melzer, S. M., Tomaskovic-Devey, D., Schunck, R., & Jacobebbinghaus, P. (2018). A Relational Inequality Approach to First- and Second-Generation Immigrant Earnings in German Workplaces. *Social Forces*, *97*(1), 91–128. https://doi.org/10.1093/sf/soy021

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and personality psychology compass*, *4*(5), 344–357.

Mincer, J. (1974). *Schooling, experience, and earnings*. National Bureau of Economic Research; distributed by Columbia University Press.

Mize, T. D. (2024). Competence Perceptions at the Intersection of Gender and Race-Ethnicity. *Social Psychology Quarterly*, 01902725241289875. https://doi.org/10.1177/01902725241289875

Monk, E. P. (2022). Inequality without Groups: Contemporary Theories of Categories, Intersectional Typicality, and the Disaggregation of Difference. *Sociological Theory*, *40*(1), 3–27. https://doi.org/10.1177/07352751221076863

Monk, E. P., Esposito, M. H., & Lee, H. (2021). Beholding Inequality: Race, Gender, and Returns to Physical Attractiveness in the United States. *American Journal of Sociology*, *127*(1), 194–241. https://doi.org/10.1086/715141

Morrow, P. C., McElroy, J. C., Stamper, B. G., & Wilson, M. A. (1990). The Effects of Physical Attractiveness and Other Demographic Characteristics on Promotion Decisions. *Journal of Management*, *16*(4), 723–736. https://doi.org/10.1177/014920639001600405

Nault, K. A., Pitesa, M., & Thau, S. (2020). The Attractiveness Advantage At Work: A Cross-Disciplinary Integrative Review. *Academy of Management Annals*, *14*(2), 1103–1139. https://doi.org/10.5465/annals.2018.0134

Neilson, W., & Ying, S. (2016). From taste-based to statistical discrimination. *Journal of Economic Behavior & Organization*, *129*, 116–128. https://doi.org/10.1016/j.jebo.2016.06.001

Neuberg, S. L., & Sng, O. (2013). A Life History Theory of Social Perception: Stereotyping at the Intersections of Age, Sex, Ecology (and Race). *Social Cognition*, *31*(6), 696–711. https://doi.org/10.1521/soco.2013.31.6.696

Nicholson, S. P., Coe, C. M., Emory, J., & Song, A. V. (2016). The Politics of Beauty: The Effects of Partisan Bias on Physical Attractiveness. *Political Behavior*, *38*(4), 883–898. https://doi.org/10.1007/s11109-016-9339-7

Norenzayan, A. (2013). *Big gods: How religion transformed cooperation and conflict*. Princeton University Press.

Norenzayan, A. (2014). Does religion make people moral? *Behaviour*, *151*(2-3), 365–384. https://doi.org/10.1163/1568539X-00003139

Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science*, *322*(5898), 58–62.

Northover, S. B., Quillien, T., Conroy-Beam, D., & Cohen, A. B. (2024). Religious signaling and prosociality: A review of the literature. *Evolution and Human Behavior*, *45*(5), 106593. https://doi.org/10.1016/j.evolhumbehav.2024.06.002

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*(2), 315–324. https://doi.org/10.1016/j.jesp.2009.12.002

Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, *112*(3), 536–552. https://doi.org/10.1037/0033-2909.112.3.536

Pandey, G., & Zayas, V. (2021). What is a face worth? Facial attractiveness biases experience-based monetary decision-making. *British Journal of Psychology*, *112*(4), 934–963. https://doi.org/10.1111/bjop.12495

Parks, F. R., & Kennedy, J. H. (2007). The Impact of Race, Physical Attractiveness, and Gender On Education Majors' and Teachers' Perceptions of Student Competence. *Journal of Black Studies*, *37*(6), 936–943. https://doi.org/10.1177/0021934705285955

Patel, D. S. (2012). Concealing to reveal: The informational role of Islamic dress. *Rationality and Society*, *24*(3), 295–323. https://doi.org/10.1177/1043463112440683

Patzer, G. L. (1985). *The Physical Attractiveness Phenomena*. Springer US. https://doi.org/10.1007/978-1-4757-0202-6

Pazhoohi, F., & Hosseinchari, M. (2014). Effects of Religious Veiling on Muslim Men's Attractiveness Ratings of Muslim Women. *Archives of Sexual Behavior*, *43*(6), 1083–1086. https://doi.org/10.1007/s10508-014-0259-5

Pedulla, D. S. (2014). The Positive Consequences of Negative Stereotypes: Race, Sexual Orientation, and the Job Application Process. *Social Psychology Quarterly*, *77*(1), 75–94. https://doi.org/10.1177/0190272513506229

Pedulla, D. S. (2018). How Race and Unemployment Shape Labor Market Opportunities: Additive, Amplified, or Muted Effects? *Social Forces*, *96*(4), 1477–1506. https://doi.org/10.1093/sf/soy002

Penner, A. M., & Saperstein, A. (2013). Engendering Racial Perceptions: An Intersectional Analysis of How Social Status Shapes Race. *Gender & Society*, *27*(3), 319–344. https://doi.org/10.1177/0891243213480262

Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review*, *62*(4), 659–661.

Piazzalunga, D. (2015). Is There a Double-Negative Effect? Gender and Ethnic Wage Differentials in Italy. *LABOUR*, *29*(3), 243–269. https://doi.org/10.1111/labr.12052

Platow, M. J., Foddy, M., Yamagishi, T., Lim, L., & Chow, A. (2012). Two experimental tests of trust in in-group strangers: The moderating role of common knowledge of group membership. *European Journal of Social Psychology*, *42*(1), 30–35.

Polavieja, J. G., Lancee, B., Ramos, M., Veit, S., & Yemane, R. (2023). In your face: A comparative field experiment on racial discrimination in Europe. *Socio-Economic Review*, *21*(3), 1551–1578. https://doi.org/10.1093/ser/mwad009

Power, E. A. (2017a). Social support networks and religiosity in rural South India. *Nature Human Behaviour*, *1*(3), 0057.

Power, E. A. (2017b). Discerning devotion: Testing the signaling theory of religion. *Evolution and Human Behavior*, *38*(1), 82–91. https://doi.org/10.1016/j.evolhumbehav.2016.07.003

Przepiorka, W., & Berger, J. (2017, September). Signaling Theory Evolving: Signals and Signs of Trustworthiness in Social Exchange. In B. Jann & W. Przepiorka (Eds.), *Social dilemmas, institutions, and the evolution of cooperation* (pp. 373–392). De Gruyter Oldenbourg. https://doi.org/10.1515/9783110472974-018

Purzycki, B. G., Apicella, C., Atkinson, Q. D., Cohen, E., McNamara, R. A., Willard, A. K., Xygalatas, D., Norenzayan, A., & Henrich, J. (2016). Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature*, *530*(7590), 327–330.

Quillian, L., Heath, A., Pager, D., Midtbøen, A. H., Fleischmann, F., & Hexel, O. (2019). Do Some Countries Discriminate More than Others? Evidence from 97 Field Experiments of Racial Discrimination in Hiring. *Sociological Science*, *6*, 467–496. https://doi.org/10.15195/v6.a18

Quillian, L., & Lee, J. J. (2023). Trends in racial and ethnic discrimination in hiring in six Western countries. *Proceedings of the National Academy of Sciences*, *120*(6), e2212875120. https://doi.org/10.1073/pnas.2212875120

Quillian, L., Lee, J. J., & Oliver, M. (2020). Evidence from Field Experiments in Hiring Shows Substantial Additional Racial Discrimination after the Callback. *Social Forces*, *99*(2), 732–759. https://doi.org/10.1093/sf/soaa026

Quillian, L., & Midtbøen, A. H. (2021). Comparative Perspectives on Racial Discrimination in Hiring: The Rise of Field Experiments. *Annual Review of Sociology*, *47*, 391–415. https://doi.org/10.1146/annurev-soc-090420-035144

Ranta, M. (2017). Master Narratives and the Pictorial Construction of Otherness: Anti-Semitic Images in the Third Reich and Beyond. *Contemporary Aesthetics*, *15*(1). https://digitalcommons.risd.edu/liberalarts_contempaesthetics/vol15/iss1/25

Ranzini, G., & Rosenbaum, J. E. (2020). It's a match (?): Tinder usage and attitudes toward interracial dating. *Communication Research Reports*, *37*(1-2), 44–54. https://doi.org/10.1080/08824096.2020.1748001

Reece, R. L. (2019). Coloring Weight Stigma: On Race, Colorism, Weight Stigma, and the Failure of Additive Intersectionality. *Sociology of Race and Ethnicity*, *5*(3), 388–400. https://doi.org/10.1177/2332649218795185

Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evolution and Human Behavior*, *26*(2), 186–201. https://doi.org/10.1016/j.evolhumbehav.2004.08.014

Ridgeway, C. (1991). The Social Construction of Status Value: Gender and Other Nominal Characteristics. *Social Forces*, *70*(2), 367–386. https://doi.org/10.1093/sf/70.2.367

Ridgeway, C. L., & Kricheli-Katz, T. (2013). Intersecting Cultural Beliefs in Social Relations: Gender, Race, and Class Binds and Freedoms. *Gender & Society*, *27*(3), 294–318. https://doi.org/10.1177/0891243213479445

Ritchie, K. L., Palermo, R., & Rhodes, G. (2017). Forming impressions of facial attractiveness is mandatory. *Scientific Reports*, *7*(1), 469. https://doi.org/10.1038/s41598-017-00526-9

Robbins, B. G. (2017). Status, identity, and ability in the formation of trust. *Rationality and Society*, *29*(4), 408–448. https://doi.org/10.1177/1043463117734179

Robbins, B. G. (2023). Valid and Reliable Measures of Generalized Trust: Evidence from a Nationally Representative Survey and Behavioral Experiment. *Socius*, *9*, 23780231231192841. https://doi.org/10.1177/23780231231192841

Romano, A., Balliet, D., Yamagishi, T., & Liu, J. H. (2017). Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences*, *114*(48), 12702–12707. https://doi.org/10.1073/pnas.1712921114

Rooth, D.-O. (2009). Obesity, Attractiveness, and Differential Treatment in Hiring: A Field Experiment. *Journal of Human Resources*, *44*(3), 710–735. https://doi.org/10.3368/jhr.44.3.710

Rosar, U., Klein, M., & Beckers, T. (2008). The frog pond beauty contest: Physical attractiveness and electoral success of the constituency candidates at the North Rhine-Westphalia state election of 2005. *European Journal of Political Research*, *47*(1), 64–79. https://doi.org/10.1111/j.1475-6765.2007.00720.x

Ross, C. E., & Mirowsky, J. (2006). Sex differences in the effect of education on depression: Resource multiplication or resource substitution? *Social Science & Medicine*, *63*(5), 1400–1413. https://doi.org/10.1016/j.socscimed.2006.03.013

Rudman, L. A., & McLean, M. C. (2016). The role of appearance stigma in implicit racial ingroup bias. *Group Processes & Intergroup Relations*, *19*(3), 374–393. https://doi.org/10.1177/1368430215583152

Ruffle, B. J., & Shtudiner, Z. (2015). Are Good-Looking People More Employable? *Management Science*, *61*(8), 1760–1776. https://doi.org/10.1287/mnsc.2014.1927

Rule, N. O., & Sutherland, S. L. (2017). Social Categorization From Faces: Evidence From Obvious and Ambiguous Groups. *Current Directions in Psychological Science*, *26*(3), 231–236. https://doi.org/10.1177/0963721417697970

Ryabov, I. (2019). How Much Does Physical Attractiveness Matter for Blacks? Linking Skin Color, Physical Attractiveness, and Black Status Attainment. *Race and Social Problems*, *11*(1), 68–79. https://doi.org/10.1007/s12552-018-9245-1

Saktanber, A., & Çorbacioğlu, G. (2008). Veiling and Headscarf-Skepticism in Turkey. *Social Politics: International Studies in Gender, State & Society*, *15*(4), 514–538. https://doi.org/10.1093/sp/jxn018

Salgado, M., Núñez, J., & Mackenna, B. (2021). Expectations of trustworthiness in cross-status interactions. *Social Science Research*, *99*, 102596. https://doi.org/10.1016/j.ssresearch.2021.102596

Samples, S. T. (2019). Black Is Not Beautiful: The German Myth of Race. In P. Essed, K. Farquharson, K. Pillay, & E. J. White (Eds.), *Relating Worlds of Racism: Dehumanisation, Belonging, and the Normativity of European Whiteness* (pp. 223–244). Springer International Publishing. https://doi.org/10.1007/978-3-319-78990-3_9

Saribay, S. A., Biten, A. F., Meral, E. O., Aldan, P., Třebický, V., & Kleisner, K. (2018). The Bogazici face database: Standardized photographs of Turkish faces with supporting materials. *PLOS ONE*, *13*(2), e0192018. https://doi.org/10.1371/journal.pone.0192018

Sarpila, O., Koivula, A., & Kukkonen, I. (2024). Appearance and Social Inequalities: Physical Attractiveness as a Part of Occupation-Congruent Appearance. *Sociology*, *58*(6), 1378–1398. https://doi.org/10.1177/00380385241249038

Schaub, M., Gereke, J., & Baldassarri, D. (2020). Does Poverty Undermine Cooperation in Multiethnic Settings? Evidence from a Cooperative Investment Experiment. *Journal of Experimental Political Science*, *7*(1), 27–40. https://doi.org/10.1017/XPS.2019.19

Schick, A., & Steckel, R. H. (2015). Height, Human Capital, and Earnings: The Contributions of Cognitive and Noncognitive Ability. *Journal of Human Capital*, *9*(1), 94–115. https://doi.org/10.1086/679675

Schilke, O., Reimann, M., & Cook, K. S. (2021). Trust in Social Relations. *Annual Review of Sociology*, *47*, 239–259. https://doi.org/10.1146/annurev-soc-082120-082850

Schmaus, M., & Kristen, C. (2022). Foreign Accents in the Early Hiring Process: A Field Experiment on Accent-Related Ethnic Discrimination in Germany. *International Migration Review*, *56*(2), 562–593. https://doi.org/10.1177/01979183211042004

Schmid, I., Witkower, Z., Götz, F. M., & Stieger, S. (2022). Registered report: Social face evaluation: Ethnicity-specific differences in the judgement of trustworthiness of faces and facial parts. *Scientific Reports*, *12*(1), 18311. https://doi.org/10.1038/s41598-022-22709-9

Schneider-Strawczynski, S., & Valette, J. (2025). Media coverage of immigration and the polarization of attitudes. *American Economic Journal: Applied Economics*, *17*(1), 337–68. https://doi.org/10.1257/app.20230300

Scholaske, L. (2024). Role of anti-Asian racism for subjective well-being among people of Asian origin in Germany. *Cultural Diversity & Ethnic Minority Psychology*, *30*(2), 247–259. https://doi.org/10.1037/cdp0000557

Scholz, J. K., & Sicinski, K. (2015). Facial Attractiveness and Lifetime Earnings: Evidence from a Cohort Study. *The Review of Economics and Statistics*, *97*(1), 14–28. https://doi.org/10.1162/REST_a_00435

Schunck, R. (2016). Reich und schön? Eine Untersuchung zur ungleichen Verteilung physischer Attraktivität. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *68*(1), 113–137. https://doi.org/10.1007/s11577-015-0354-1

Schwartzman, E., & Rule, N. O. (2024). Sexual orientation as a contextual frame for attractiveness judgments. *Journal of Experimental Psychology: General*, *153*(11), 2899–2917. https://doi.org/10.1037/xge0001563

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C. M., van Assen, M. A. L. M., Liu, Y., Althoff, T., Heer, J., Kale, A., Mohamed, Z., Amireh, H., Venkatesh Prasad, V., Bernstein, A., Robinson, E., Snellman, K., Amy Sommer, S., Otner, S. M. G., Robinson, D., . . . Luis Uhlmann, E. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes*, *165*, 228–249. https://doi.org/10.1016/j.obhdp.2021.02.003

Shapir, O. M., & Shtudiner, Z. (2022). Beauty is in the eye of the employer: Labor market discrimination of accountants. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.928451

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*(9), 803–809.

Shaver, J. H., Lang, M., Krátkỳ, J., Klocová, E. K., Kundt, R., & Xygalatas, D. (2018). The boundaries of trust: Cross-religious and cross-ethnic field experiments in Mauritius. *Evolutionary Psychology*, *16*(4), 1474704918817644.

Sheen, M., Yekani, H. A. K., & Jordan, T. R. (2018). Investigating the effect of wearing the hijab: Perception of facial attractiveness by Emirati Muslim women living in their native Muslim country. *PLOS ONE*, *13*(10), e0199537. https://doi.org/10.1371/journal.pone.0199537

Sides, J., & Citrin, J. (2007). European Opinion About Immigration: The Role of Identities, Interests and Information. *British Journal of Political Science*, *37*(3), 477–504. https://doi.org/10.1017/S0007123407000257

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*(11), 1208–1214. https://doi.org/https://doi.org/10.1038/s41562-020-0912-z

Simpson, B., Harrell, A., & Willer, R. (2013). Hidden paths from morality to cooperation: Moral judgments promote trust and trustworthiness. *Social Forces*, *91*(4), 1529–1548. https://doi.org/10.1093/sf/sot015

Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H. J., & Todorov, A. (2017). For Your Local Eyes Only: Culture-Specific Face Typicality Influences Perceptions of Trustworthiness. *Perception*, *46*(8), 914–928. https://doi.org/10.1177/0301006617691786

Soroka, S. N., Helliwell, J. F., & Johnston, R. (2007). Measuring and Modelling Interpersonal Trust. In R. Johnston & F. Kay (Eds.), *Social Capital, Diversity, and the Welfare State* (pp. 95–132). UBC Press.

Sosis, R. (2005). Does Religion Promote Trust?: The Role of Signaling, Reputation, and Punishment. *Interdisciplinary Journal of Research on Religion*, *1*.

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American economic review*, *92*(3), 434–459.

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, *108*(19), 7710–7715.

StataCorp. (2021). Stata Statistical Software: Release 17.

Stavrova, O., & Siegers, P. (2014). Religious prosociality and morality across cultures: How social enforcement of religion shapes the effects of personal religiosity on prosocial and moral attitudes and behaviors. *Personality and Social Psychology Bulletin*, *40*(3), 315–333.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Suo, Z., Guo, Q., & Shen, J. (2024). Revisiting the effect of trustworthy face and attractive appearance on trust and trustworthiness behavior. *Applied Economics*, *0*(0), 1–21. https://doi.org/10.1080/00036846.2024.2316119

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *Social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks-Cole.

Tan, J. H. W., & Vogel, C. (2008). Religion and trust: An experimental study. *Journal of Economic Psychology*, *29*(6), 832–848. https://doi.org/10.1016/j.joep.2008.03.002

Tanis, M., & Postmes, T. (2005). A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology*, *35*(3), 413–424. https://doi.org/10.1002/ejsp.256

Tews, M. J., Stafford, K., & Zhu, J. (2009). Beauty Revisited: The impact of attractiveness, ability, and personality in the assessment of employment suitability. *International Journal of Selection and Assessment*, *17*(1), 92–100. https://doi.org/10.1111/j.1468-2389.2009.00454.x

Thijssen, L., Coenders, M., & Lancee, B. (2021a). Is there evidence for statistical discrimination against ethnic minorities in hiring? Evidence from a cross-national field experiment. *Social Science Research*, *93*, 102482. https://doi.org/10.1016/j.ssresearch.2020.102482

Thijssen, L., Coenders, M., & Lancee, B. (2021b). Ethnic Discrimination in the Dutch Labor Market: Differences Between Ethnic Minority Groups and the Role of Personal Information About Job Applicants—Evidence from a Field Experiment. *Journal of International Migration and Integration*, *22*(3), 1125–1150. https://doi.org/10.1007/s12134-020-00795-w

Thijssen, L., Lancee, B., Veit, S., & Yemane, R. (2021). Discrimination against Turkish minorities in Germany and the Netherlands: Field experimental evidence on the effect of diagnostic information on labour market outcomes. *Journal of*

Ethnic and Migration Studies, *47*(6), 1222–1239. https://doi.org/10.1080/1369183X.2019.1622793

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, *3*(12), 452–460. https://doi.org/10.1016/S1364-6613(99)01403-5

Thunström, L., Jones Ritten, C., Bastian, C., Minton, E., & Zhappassova, D. (2021). Trust and Trustworthiness of Christians, Muslims, and Atheists/Agnostics in the United States. *Journal for the Scientific Study of Religion*, *60*(1), 147–179. https://doi.org/https://doi.org/10.1111/jssr.12692

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton University Press.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, *66*, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Tseligka, E. D. (2016). Becoming the Other – stereotyping of Greeks by the German press. *Continuum*, *30*(6), 627–635. https://doi.org/10.1080/10304312.2016.1186604

Turner, J. C. (1975). Social comparison and social identity: Some prospects for intergroup behaviour. *European Journal of Social Psychology*, *5*(1), 1–34. https://doi.org/10.1002/ejsp.2420050102

Umberson, D., & Hughes, M. (1987). The Impact of Physical Attractiveness on Achievement and Psychological Well-Being. *Social Psychology Quarterly*, *50*(3), 227–236. https://doi.org/10.2307/2786823

Valfort, M.-A. (2020). Anti-Muslim discrimination in France: Evidence from a field experiment. *World Development*, *135*, 105022. https://doi.org/10.1016/j.worlddev.2020.105022

Valmori, A., Meral, E. O., Hale, M.-L., Rusconi, P., & Brambilla, M. (2023). On the influence of implicit race attitudes on explicit trustworthiness judgments: An investigation of the perceivers and targets' race and gender intersection. *Visual Cognition*, *31*(8), 584–598. https://doi.org/10.1080/13506285.2024.2315800

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803. https://doi.org/10.1016/j.cognition.2008.07.002

van Tubergen, F. (2025). Perceived Societal Discrimination among Immigrants across 17 European Countries: Extending the Integration Paradox? *European Societies*, 1–33. https://doi.org/10.1162/euso_a_00004

Veit, S., Arnu, H., Di Stasio, V., Yemane, R., & Coenders, M. (2022). The "Big Two" in Hiring Discrimination: Evidence From a Cross-National Field Experiment. *Personality and Social Psychology Bulletin*, *48*(2), 167–182. https://doi.org/10.1177/0146167220982900

Veit, S., & Essien, I. (2022). *Faces. Version: 1.0.0* (tech. rep.). Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM). https://doi.org/10.34882/dezim.dpd-fac.f.1.0.0

Veit, S., & Thijsen, L. (2021). Almost identical but still treated differently: Hiring discrimination against foreign-born and domestic-born minorities. *Journal of Ethnic and Migration Studies*, *47*(6), 1285–1304. https://doi.org/10.1080/1369183X.2019.1622825

Vernby, K., & Dancygier, R. (2019). Can immigrants counteract employer discrimination? A factorial field experiment reveals the immutability of ethnic hierarchies. *PLOS ONE*, *14*(7), e0218044. https://doi.org/10.1371/journal.pone.0218044

Vuolo, M., Uggen, C., & Lageson, S. (2018). To Match or Not to Match? Statistical and Substantive Considerations in Audit Design and Analysis. In S. M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 119–140). Springer International Publishing. https://doi.org/10.1007/978-3-319-71153-9_6

Walzenbach, S. (2019). Hiding Sensitive Topics by Design? An Experiment on the Reduction of Social Desirability Bias in Factorial Surveys. *Survey Research Methods*, *13*(1), 103–121. https://doi.org/10.18148/srm/2019.v1i1.7243

Webster, M., & Driskell, J. E. (1978). Status Generalization: A Review and Some New Data. *American Sociological Review*, *43*(2), 220–236. https://doi.org/10.2307/2094700

Webster, M., & Driskell, J. E. (1983). Beauty as Status. *American Journal of Sociology*, *89*(1), 140–165.

Weichselbaumer, D. (2020). Multiple Discrimination against Female Immigrants Wearing Headscarves. *ILR Review*, *73*(3), 600–627. https://doi.org/10.1177/0019793919875707

Wiemers, S. A., Stasio, V. D., & Veit, S. (2024). Stereotypes about muslims in the netherlands: An intersectional approach. *Social Psychology Quarterly*, *87*(4), 440–460. https://doi.org/10.1177/01902725231219688

Wigger, I. (2019). Anti-Muslim racism and the racialisation of sexual violence: 'intersectional stereotyping' in mass media representations of male Muslim migrants in Germany. *Culture and Religion*, *20*(3), 248–271. https://doi.org/10.1080/14755610.2019.1658609

Wilkins, C. L., Chan, J. F., & Kaiser, C. R. (2011). Racial stereotypes and interracial attraction: Phenotypic prototypicality and perceived attractiveness of Asians. *Cultural Diversity and Ethnic Minority Psychology*, *17*(4), 427–431. https://doi.org/10.1037/a0024733

Williams, D. (2024). Racism without 'race': Colorblindness, blackness and everyday racism in contemporary Germany. *Identities*, 1–18. https://doi.org/10.1080/1070289X.2024.2388961

Willis, J., & Todorov, A. (2006). First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, R. K., & Eckel, C. C. (2006). Judging a Book by its Cover: Beauty and Expectations in the Trust Game. *Political Research Quarterly*, *59*(2), 189–202. https://doi.org/10.1177/106591290605900202

Wilson, R. K., & Eckel, C. C. (2011). Trust and social exchange. *Cambridge handbook of experimental political science*, 243–57.

Wong, J. S., & Penner, A. M. (2016). Gender and the returns to attractiveness. *Research in Social Stratification and Mobility*, *44*, 113–123. https://doi.org/10.1016/j.rssm.2016.04.002

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Cengage Learning.

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, *117*(2), 364–385. https://doi.org/10.1037/pspi0000160

Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of theoretical Biology*, *53*(1), 205–214.

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking Smart and Looking Good: Facial Cues to Intelligence and their Origins. *Personality and Social Psychology Bulletin*, *28*(2), 238–249. https://doi.org/10.1177/0146167202282009

Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L. A., & Montepare, J. M. (2015). Faces and first impressions. In *APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition* (pp. 251–276). American Psychological Association. https://doi.org/10.1037/14341-008

Zebrowitz, L. A., Montepare, J. M., & Lee, H. K. (1993). They don't all look alike: Individual impressions of other racial groups. *Journal of Personality and Social Psychology*, *65*(1), 85–101. https://doi.org/10.1037/0022-3514.65.1.85

Zhang, N., Aidenberger, A., Rauhut, H., & Winter, F. (2019). Prosocial behaviour in interethnic encounters: Evidence from a field experiment with high-and low-status immigrants. *European Sociological Review*, *35*(4), 582–597.

Zhang, N., Gereke, J., & Baldassarri, D. (2022). Everyday Discrimination in Public Spaces: A Field Experiment in the Milan Metro. *European Sociological Review*, *38*(5), 679–693. https://doi.org/10.1093/esr/jcac008

Zhao, N., Zhou, M., Shi, Y., & Zhang, J. (2015). Face Attractiveness in Building Trust: Evidence from Measurement of Implicit and Explicit Responses. *Social Behavior and Personality: an international journal*, *43*(5), 855–866. https://doi.org/10.2224/sbp.2015.43.5.855

Ziller, C. (2022). Established and Excluded? Immigrants' Economic Progress, Attitudes toward Immigrants, and the Conditioning Role of Egalitarianism and Intergroup Contact. *Political Studies*, *70*(1), 236–256. https://doi.org/10.1177/0032321720953561

Zolberg, A. R., & Woon, L. L. (1999). Why Islam is like Spanish: Cultural incorporation in Europe and the United States. *Politics & Society*, *27*(1), 5–38.

Zschirnt, E. (2019). Research Ethics in Correspondence Testing: An Update. *Research Ethics*, *15*(2), 1–21. https://doi.org/10.1177/1747016118820497

Zschirnt, E. (2020). Evidence of Hiring Discrimination Against the Second Generation: Results from a Correspondence Test in the Swiss Labour Market. *Journal of International Migration and Integration*, *21*(2), 563–585. https://doi.org/10.1007/s12134-019-00664-1

Zschirnt, 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. https://doi.org/10.1080/1369183X.2015.1133279