Beauty Contest Revisited: The Effects of Perceived Attractiveness, Competence, and Likability on the Electoral Success of German MPs

Jäckle, Sebastian; Metz, Thomas

Postprint / Postprint
Zeitschriftenartikel / journal article

Mitglied der
Leibniz-Gemeinschaft

gesis
Leibniz-Institut
für Sozialwissenschaften

Beauty contest revisited – the effects of perceived attractiveness, competence and likability on the electoral success of German MPs

Sebastian Jäckle, University of Freiburg

Thomas Metz, University of Freiburg

Abstract

We test the effects of physical appearance on electoral outcomes for the 2013 German national elections. We find that a candidate's perceived attractiveness and to a lesser extent competence vis-á-vis his or her closest contestant increases chances of winning a direct mandate, while likability plays a minor role. Additionally we find the appearance effects to be conditioned by gender, age and incumbency status. Our study advances existing research in four ways: First, we capture relative differences in appearances, which resembles real-world situations more closely than absolute measures. Second, we proceed beyond a one-dimensional assessment of appearance by simultaneously analysing attractiveness, competence, and likability, including interactions. Third, the central role of parties in the German mixed electoral system makes an especially tough test for appearance based effects. Fourth, we use rater response latency to weight our measurement with an assessment of ambivalence.

1. Introduction

"What is beautiful is good" – at least for most of the time and most of the people.

Writing over forty years ago, Dion, Berscheid and Walster (1972) for the first time provided evidence that beautiful people are "assumed to possess more socially desirable personality traits" and are "expected to lead better lives" (Dion et al. 1972, 285). Since then, a "beauty premium" as Praino and colleagues (2014, 1097) term it, has been detected in sociological, economic and psychological studies within virtually all fields of life. Attractive babies get more affectionate care by their mothers (Langlois et al. 1995), good-looking pupils receive better grades (Dukake et al. 2012) and are more likely to obtain a college degree (Gordon et al. 2014). Attractive persons have a better chance to get a call-back when applying for a job (Bóo et al. 2013), they get paid better (Hamermesh and Biddle 1994) and they do not have to deliver the same performance as their unattractive counterparts in sports (e.g. in soccer, see Rosar et al. 2010). Still, while beauty often pays, it can also have negative effects. For example, Johnson et al. (2010) show that attractiveness[1] is detrimental for women applying for masculine sex-typed jobs in which appearance is of minor importance (e.g. prison warden). This interaction effect between gender and attractiveness has been termed the "beauty is beastly" effect (Heilman and Saruwatari 1979).

In politics, several studies confirm the effects of physical attractiveness on the likelihood of being elected. Already in the mid-1970s, Efran and Patterson (1974, 354) found a significant correlation between beauty ratings of candidates and their vote share in Canadian federal elections. Subsequent studies added several controls (e.g. incumbency status) and confirmed a general effect of beauty in political systems as diverse as Australia, Finland, Brazil or Mexico

---

[1] The terms "beauty" and "attractiveness" are used interchangeably throughout this article.

(King and Leigh 2009; Berggren et al. 2010; Lawson et al. 2010) and also in Germany (Klein and Rosar 2005). The explanation why people seem to base their vote at least partly on candidates' appearance is simple. During campaigns, pictures of the candidates are in most cases readily available (e.g. from campaign posters or newspapers). The electorate uses these pictures as "thin slices" (Ambady and Rosenthal 1992) of information about the contestants to infer personal traits relevant to them which they cannot readily learn about otherwise. When trying to infer these attributes, candidate appearances such as attractiveness are usually easy to assess. This creates a halo effect in which the easily available perception of beauty translates to other attributes and outshines other markers one might draw on. As a consequence, beauty regularly serves as a (subconscious) basis for inferring the trait of interest. In that sense, perceived beauty works as a heuristic which enables voters who do not know much (or anything) about the candidates – and thus could not make a well-founded choice – to gather enough information for deciding whom they want to vote for. However, voters may also resort to visual features when information relevant for a rational and deliberative decision – e.g. information given in a newspaper article about the candidate – is readily available (Barrett and Barrington 2005).[2]

Obviously, other cues about candidates are available too, such as race, gender (McDermott 2009), or party (e.g. Dancey and Sheagle 2013). Here, the decline in partisanship in most Western countries (Dalton 2000; Arzheimer 2006) has raised expectations that candidates – and, subsequently, their appearances – might become increasingly relevant for vote choice.

---

[2] The usual perspective on heuristics is that they serve as a cheap but somewhat inaccurate alternative to the inefficient processing of large amounts of information. Yet, political psychologists have rejected this idea already in the late 1960s and contended that voters ignore more accurate information even if it was readily available. Quoting Sears, Bull and Hawkes (1982, 95) put this strikingly: "It is easier to base one's decision on how a person looks rather than on the arguments he is putting forward." Sears even took it one step further when he argued that "the primary purpose of raising issues at all may be simply to provide something for the candidate to talk about" (Sears 1969, 368). In his position it is therefore not primarily policy positions which candidates want to get into the newspapers, but their faces.

Together with other aspects such as a growing media-focus on persons instead of parties, the possibility that individuals may themselves serve as cues for voters has led to the concern that politics could undergo a process of *personalization*. In such a situation, strong effects from candidate appearances would be a rather mixed blessing. While the literature for Germany so far agrees that although candidates in general do matter, support for a *personalization* is rather meagre and that candidate effects do not change much over time (Pappi and Shikano 2001; Ohr and Klein 2013), the picture is not yet complete since most work focuses on chancellor candidates, neglecting the level of constituencies. Here, work by Rosar and Klein (Rosar et al. 2008; Klein and Rosar 2005) has already shown that sometimes substantial effects of physical attractiveness on electoral success exist. Yet, their work has so far mostly focused on attractiveness and not included other aspects of physical appearance such as perceived likability or inferred competence (but see Klein and Rosar 2005). Furthermore, although beauty is in part relative since the perception of attractiveness is affected by available alternatives (Kenrick and Gutierres 1980), studies on appearance effects in politics virtually all rely on absolute measures of candidate looks (modelling voter decision much akin to a jury in a beauty contest) instead of capturing attractiveness as a relative decision between two or more available candidates who do not exist independently of each other.

The studies presented above have substantially contributed to our understanding of elections and electoral outcomes by complementing standard explanations of vote choice such as sociological theorizing, the social-psychological model of voting, and especially rational choice approaches. That heuristics based on candidates' appearances may significantly affect voters' decisions not only warrants scientific inquiry in itself but also raises both very practical issues and questions of broader democratic relevance. Indeed, if we accept that it is not solely one's ideological position or one's convincing arguments that help one to get elected, our understanding of (democratic) politics would appear incomplete without taking visual impressions into account, especially in a world that increasingly relies on (social)

media with their heavy use of images for information. In this article, we contend that research on voting behaviour can only gain from more explicitly paying attention to the fact that voters' choices are also affected by extra-political factors such as the superficial looks of candidates competing in a kind of "beauty contest". At first, such a perspective may not only feel unfamiliar to political scientists given our usual theorizing based on e.g. ideological distance measures and rational choices, and indeed, it may even appear normatively undesirable. Yet, it is a factor nonetheless and the picture would just not be complete without it. As such it makes perfect sense to focus on the electoral effects of a candidate's appearance. However, once we explore the role of facial traits, an obvious question is which trait it is that helps the most to get elected. With the existing literature disagreeing about this question – some say it is attractiveness while others are more inclined towards perceived competence – we take a neutral stance and test these two most often mentioned facial traits against each other. Furthermore, we introduce likability as a third type of perception that has been neglected in most studies so far and also test it against the other two (see next paragraph for a more thorough explanation of these three traits and existing studies on their effects on voting decisions). Our research is mostly exploratory, so we refrain from formulating explicit hypotheses which of the three traits matters most. Yet, we certainly assume that candidates' appearance plays a significant and relevant role even compared to important explanatory factors such as the incumbency status. Additionally, if appearance matters, it is likely that it does not affect all candidates in the same way. Rather, we expect conditional effects between the facial traits and further characteristics of the candidate to be at work. Again, while existing studies have already indicated that different interactions may be relevant, a thorough test of these interactions is still lacking. We therefore chose three of the assumed conditioning factors most often mentioned and test whether they influence the appearance effects.

To sum up, in this article we will join and extend the line of research presented above by (1) testing whether perceived attractiveness, competence or likability of direct candidates to the

German Bundestag in the 2013 election can serve as relevant predictors for their electoral success and (if any) which trait matters most. Furthermore, we will (2) test whether these perceived appearance effects are conditioned by gender, incumbency and age.

Preceding the main analysis our study additionally tests which features of the candidates' faces, such as wearing a beard or glasses on the photos, impact on the perception of the three traits. In case one of the three traits emerges as a relevant determinant for the vote share, knowing which factors can help to boost this trait is certainly important for parties and candidates.

2. A matter of beauty, competence or likability?

While the general effect of candidates' physical appearance on their chances to be elected by now is stably anchored, there are different ideas about the underlying causal relations. On the one hand, the studies referred to above argue for a "what is beautiful is good" halo effect. They regard attractiveness as a fundamental perception about a person which subsequently colours all kinds of ascribed attributes and therefore should ultimately also affect voters' decisions. On the other hand, Todorov and colleagues (Todorov et al. 2005; Olivola and Todorov 2010) linked rapid, unreflected inferences of competence based on candidate pictures to election results of U.S. House and Senate races. According to them, attractiveness may matter to some degree for how competent voters judge a candidate (along with babyfacedness, familiarity and age [see Olivola and Todorov 2010, 92]), but in the end perceived competence is decisive. Expressed in terms of the social-psychological model of voting (Campbell et al. 1960), the difference between both perspectives is whether role-related candidate attributes (i.e. competence) or more role-distant attributes (i.e. attractiveness) deliver the decisive momentum.

Joining the line of Todorov et al., Armstrong and colleagues show that ratings of candidates' "facial competence" were better predictors for "the popular vote winners in the presidential primaries and general election [than] […e]arly polling results" (Armstrong et al. 2010, 519). According to them, parties could actively increase their electoral chances in the same way that companies merchandize their products by using an attractive packaging. Another study suggests that this is already done, particularly in competitive electoral districts: Challengers sent to "tossup" districts in U.S. House and Senate races have significantly higher facial competence scores than challengers in electorally safe districts (Atkinson et al. 2009, 233). Ballew and Todorov (2007) demonstrate that even very short exposures (100ms) to facial pictures of both candidates in a gubernatorial race suffice to predict the winner by asking who appears more competent.

Studies testing both – perceived attractiveness and competence – are scarce and so far have come to no consistent result. While Berggren et al. conclude that at least for Finnish non-incumbent candidates "beauty is more strongly correlated with success than […] perceived competence" (2010, 8), Olivola and Todorov arrive at a contrary result (2010, 95).

Additionally to attractiveness and competence, we test for likability as a third, role-unrelated predictor which so far has been treated as a stepchild in appearance studies. While likability has mostly been examined for the U.S. after Wattenberg (1992) depicted "the rise of candidate-centered politics" (see Bishin et al. 2006) one shortcoming of most studies is that they only focus on well-known politicians. Yet, if people already follow their gut feelings when assessing well-known politicians like U.S. presidential candidates (where information is easily available), likability should play an even greater role when voters are less familiar with the candidates. Additionally, appearance studies usually consider likability as a further control at best, failing to elaborate potential interactions. Thus, it comes as no surprise that these studies find only little, if any, evidence of an independent likability-effect: For example,

Berggren and colleagues find no additional effect of likability when controlling for beauty, arguing that both assessments are collinear (Berggren et al. 2010, 14). However, Rule and colleagues (2010) have shown that a perceived factor "power" (but not "warmth") was predictive for the electoral success of U.S. senators while for the Japanese diet the effects were reversed. Given that "warmth" contained perceived likability, the apparent underlying cultural variation of the effect cautions against dismissing likability for Germany based mostly on negative findings from the U.S. At the level of constituencies, no study has tested for likability so far. Analysing the effects of likability appears all the more important given that Olivola and Todorov find that likability has an effect on hypothetical but not on actual votes which could "suggest that experiments limited to the laboratory may overestimate the role of these inferences [judgements related to likability] in predicting real life outcomes" (Olivola and Todorov 2010, 94). We therefore include likability to test systematically whether and when it has an additional effect compared to attractiveness and competence at the level of constituencies.

Summing up, this leads to our first research question:

Q1: To what extent do perceived attractiveness, competence and likability impact on the election results for direct candidates to the German Bundestag? And which of these traits is most important?

Regarding the second part of this question Praino et al. (2014) have argued that at least with respect to attractiveness and competence there is no simple answer. They found that whether competence or attractiveness is more important depends on the configuration of the contestants regarding their gender: competence is more important for inter-gender races while in intra-gender races, it is attractiveness because

"[b]etween two men or two women, it is fairly easy to determine with a glance who is more attractive than the other; when comparing a man with a woman, however such assessment becomes

much trickier, even at times impossible. […] Voters tend to be easily influenced by good-looking candidates when it is easy for them to choose which candidate looks best. When such assessment becomes more complicated, voters end up choosing the second easiest path, that is, they determine who appears to be more competent" (Praino et al. 2014, 1111).

Different studies suggest that the effects of appearance, particularly of attractiveness are conditioned by other interactions:

a) Male candidates seem to benefit more from beauty than females, probably because physical attractiveness may also carry negative connotations particularly for women (the "dumb blonde syndrome" [King and Leigh 2009, 591]).

b) Attractiveness has a stronger effect on challengers than on incumbents, probably because voters know more about incumbents already and therefore need to resort less to thin slices of information on their physical characteristics for arriving at a decision (King and Leigh 2009, 592).

c) With attractiveness being correlated strongly with youth it can be argued that particularly in races where a young candidate competes against an old candidate, attractiveness should play a more relevant role (see e.g. McLellan and McKelvie 1993).

These points lead to our second research question:

Q2: To what extent are the effects of the three appearance traits conditioned by (a) the gender of the candidate, and (b) the configuration of candidates within the electoral district regarding gender, age and incumbency status.

3. Research Design

This section introduces our research design. After explaining why German federal elections provide a good setting for analysing appearance effects, we describe how we measured perceived attractiveness, competence and likability with an online survey. The third subsection explores which features in the pictures foster the perception of the three traits. We then describe our weighting procedure for ambiguity in the ratings based on latency times before subsection five describes our dependent variable and the controls.

3.1.Testing appearance based effects in elections to the German Bundestag

Many studies of the effects of perceived appearance focus on the United States where personalization is comparatively strong and parties relatively weak, making it likely to find appearance effects. If, however, a "beauty premium" is universal, we should also find it in a country where the role of parties is stronger and politics are less personalized. Germany is well suited for such a test since it has strong and disciplined parties which are the main focal points for voting decisions. Furthermore, it is large enough (in terms of electoral districts/candidates) to allow for a meaningful statistical analysis. While personalization has somewhat increased during the last decades in Germany, it is still low compared to the US, and most studies find its effects to be limited to top-politicians such as party leaders and chancellor candidates (Kaase 1994; Pappi and Shikano 2001; Brettschneider 2002). Furthermore, party identification is still strong in Germany and has a bigger influence on the final voting decision than in the US presidential system (Berglund et al. 2005, 105-106). Most students of German politics would therefore agree that it is still the parties and not candidates which shape federal elections.

Although the German electoral system is in the end a PR system, one of the two votes (*Erststimme* or "first vote") is given to a candidate (and not to a party). Thus, in each of the 299 electoral districts there is a competition not only between the parties, but also between so called "direct candidates". Due to the structure of the German party system, in most districts

only candidates from the conservative Christian-Democratic Union CDU (or its sister party CSU in the state of Bavaria) or Social Democratic SPD have realistic chances to get elected.[3] Apart from the parties' front-runners, election campaigns within the districts are also focused on the respective direct candidates to a substantial degree. In practice this means that several weeks before the election the streets are full with election posters showing images of the candidates. We can therefore safely assume that, at least for the major parties, voters know how the candidates look like.

3.2. Measuring perceived attractiveness, competence and likability

For each electoral district we collected photographs of the two candidates who won most votes in that district at the 2013 general elections. In most cases we used photos from the candidates' official websites. Sometimes we also resorted to the Bundestag's or a parliamentary party group's homepage, or to newspapers. We strived to collect pictures that resembled those used in election campaigns as much as possible since it is these pictures that voters will most probably have in mind. Therefore, if a candidate's appearance has any effect, it should be based on the pictures the public has access to. The pictures we used all showed the candidate's complete face, and usually included the shoulders. Presenting not only the facial expression (we excluded pictures where candidates showed strong facial expressions like wide smiles), but also hairstyle, clothing and jewelry makes sense because these non-facial aspects of the picture have shown to be at least equally important as the sole physiognomy (Spezio et al. 2012). We selected photos without distracting backgrounds, scaled them to the same size and transformed them to grey-scale. To prevent raters from associating the pictures with certain parties all party symbols were removed (e.g. lapel pins).

---

[3] In East German constituencies, socialist LINKE are also often on a level with CDU and SPD. The GREEN party also has some strongholds, particularly in larger cities, where they possess somehow realistic chances to win a direct constituency seat.

In our online survey we showed each rater 25 pairs of candidates, each pair representing one electoral district. Raters were asked to indicate which of the two persons they deemed to be more attractive, competent and likable.[4] Our measurement thus differs from absolute ratings. The idea of comparing only the two main alternatives draws on the observation by Kenrick and Gutierres (1980) who showed that subjects rated a person as comparatively less attractive after having seen pictures of beautiful people – apparently, when it comes to appearances, alternatives do matter. We therefore expect our approach to reproduce the situation in an electoral district more faithfully. First, real-life voters are probably unaware of the looks of most candidates outside their district. Thus presenting raters with alternatives that may affect their rating but that real-life voters probably never see may at best add noise. Instead, constraining measurement to alternatives effectively available to voters should minimize differences between voters and raters. Second, virtually all models of choice reason comparatively by assuming that an option is not picked because of its absolute merit but rather because it is the best (or closest) available alternative. Any effects of perceived personal traits should thus be based on the relative advantages or disadvantages a candidate has compared to his or her contestant within the electoral district. And third, to the extent that rating appearances is a matter of cognitive effort, a relative model is more parsimonious since it avoids an absolute standard of reference. Our measurement takes these points into account and follows the suggestions by Ballew and Todorov (2007) as well as by King and Leigh (2009) who also have stressed the importance of measuring perceived traits in a relative way.

While relative judgments may be more accurate, our measurement is, of course, no silver bullet since rating only the two main candidates leaves open the possibility that a third

---

[4] With our study being interested in the perception of the candidates appearance it was not necessary to define the three traits in more detail. Every rater had the possibility of applying his or her own understanding of attractiveness, competence and likability. In the end we define the three traits by using the empirical data. I.e. a person is more attractive (competent/likable) compared to his or her contestant if raters had chosen him or her more often as the more attractive (competent/likable) candidate.

candidate might also affect voters' assessments. However, the problem is probably not overly severe: First, it is usually the two top candidates that set the stage for the contest since only they have – at least in most cases – realistic chances of winning.[5] And second, even if effects from a missing third candidate enter real-life judgments, there is no reason to assume that the effects are systematic and thus more than noise.

All studies on the effects of personal traits face a common problem: When raters recognize the candidates, their ratings might no longer rely on appearance alone but also depend on the candidate's person or party. Different strategies have been adopted as solutions (see Praino et al. 2014, 1102). Little and colleagues (2007) use only the shapes of the candidates' faces which makes it impossible for participants to identify whom they rate. Others manipulate the photos to achieve the same effect (Armstrong et al., 2010; Todorov et al., 2005). Both approaches are not helpful for our purpose since they undermine the idea of rating candidates based on publicly used pictures. A third strategy is to make plausible that raters do not know the candidates. For this purpose Lawson and his colleagues (2010) draw on US-American and Indian students to rate politicians from Mexico and Brazil. Antonakis and Dalgas (2009) present children and adults from Switzerland with photographs of candidates for the French National Assembly. Asked which person they would prefer as a captain on a boat journey the children's answers were as good as the adult's in predicting who actually won in the election.

We take a different, two-stage approach. First, we excluded well-known politicians from the sample with a pre-test.[6] Removing these individuals is both a precondition for obtaining

---

[5] In more than 90 per cent of the electoral districts, the party with the second most first votes has a lead of at least five percentage points over the third-placed party. Concentrating on the two strongest parties in each district is therefore justifiable in most cases.

[6] In this pre-test, eleven students from the University of Heidelberg and six from the University of Freiburg took part in an online survey in which they indicated for each of the 598 candidates whether they recognized or were able to associate him or her with a party. We excluded an electoral district if more than two pre-testers recognized one or both candidates. The pre-testers did not participate in the main survey.

ratings based on the presented photos alone and can also be justified with respect to the underlying voter model: For well-known politicians (e.g. chancellor Angela Merkel), voters can draw on much more information than for the average candidate, requiring them to rely less on information shortcuts, subsequently decreasing appearance based effects. On the basis of the pre-test we excluded 40 of the 299 electoral districts. As a second step to assure unbiased ratings, we also asked the raters at the end of the main survey whether they had recognized a person and whether they knew his or her name. Districts in which a rater recognized a candidate were then excluded from the analysis for that rater.[7]

---

[7] Of the 11,262 pairs of candidates presented, participants recognized a politician in only 48 instances. These instances involved 31 different politicians in total which implies that no politician was recognized systematically, suggesting that the pre-test worked as intended.

Figure 1: Screenshot of Online Rating Tool
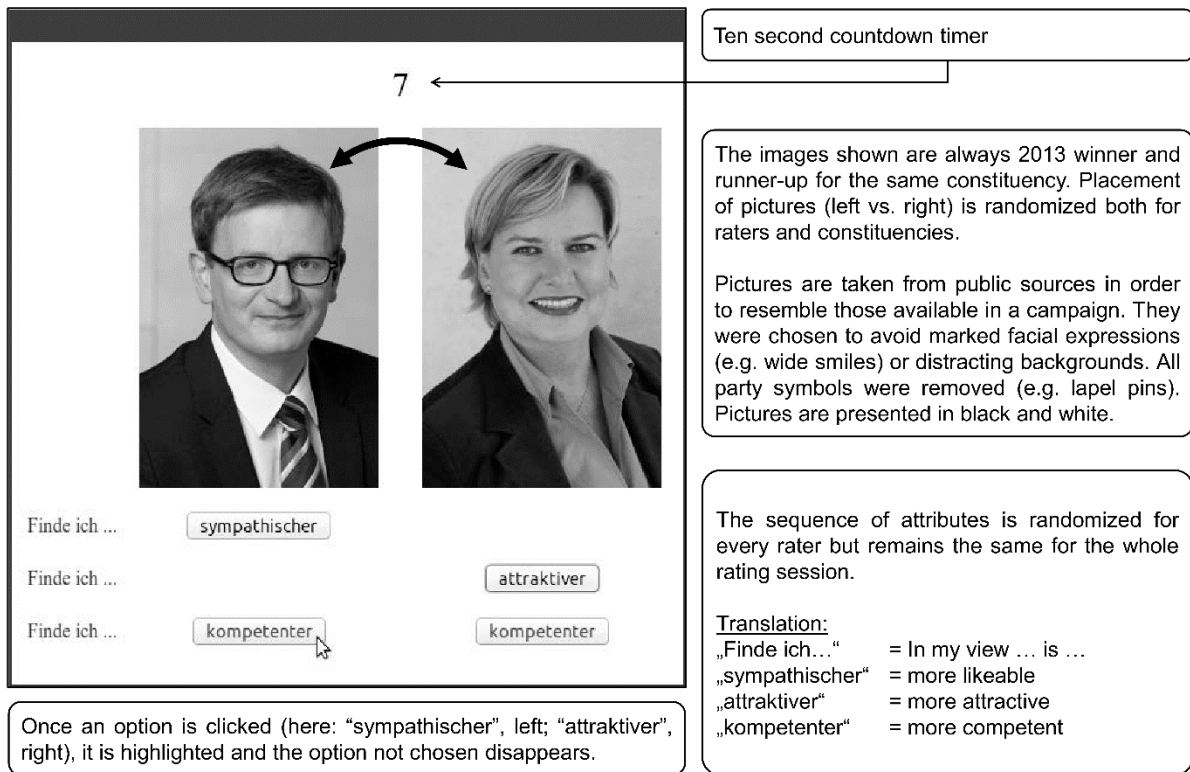
Raters were unaware that they rated politicians. Instead, they were given a cover story emphasizing that they participated in a survey about the correlation between appearance and social cooperation.[8] Candidate pairs presented were chosen randomly from the entirety of all 259 included constituencies.[9] We randomized which candidate appeared on the left/right half of the screen and (across raters) the ordering of buttons for rating. A screenshot of the rating tool can be seen in figure 1.

---

[8] Nevertheless, even if they did not know the persons, it can not be completely ruled out that participants realized that they in fact rated politicians, as the pictures clearly resemble headshots of politicians like those often visible in public.

[9] This makes a total of 518 candidates for whom we collected the pictures. 364 of them were male and 154 female. In 123 districts two male candidates ran against each other, in 18 districts it was two females and in 118 a man ran against a woman. In addition, male candidates have a more than two and a half times higher chance of winning when they compete against a female candidate (85 to 33). The average candidate was born in 1963 and thus was 50 years old in the election year. Ten candidates had a migration background (most of them were of Turkish descent). Yet, for none of them the greyscale pictures allowed to trace back their non-German descent in comparison to the candidates with no migration background. A systematic racial bias can therefore be ruled out.

Each rater was presented 25 pairs of pictures. While other studies had participants rate up to several hundred pairs (see e.g. Atkinson et al. 2009, 232), we chose this relatively small number to help raters remain concentrated. Furthermore, with only 25 rated pairs of candidates, raters should be less likely to develop systematic response patterns which could bias the results. Each pair was shown for ten seconds. A visually prominent countdown indicated how much time was left to click. When time was over or the participant had clicked the third item, there was a one second blank screen before the next pair of pictures was presented. In total, 449 student raters completed the survey.[10] On average each electoral district was presented to 43.48 (sd = 6.55) raters. For attractiveness, the mean number of ratings for each pair of candidates is 38.56 (sd = 6.21; min = 23; max = 61), for competence it is 40.00 (sd = 6.30; min = 22; max = 62) and for likability it is 40.52 (sd = 6.37; min = 24; max = 63) respectively. All ratings are thus based on a sufficiently large sample according to the truth of consensus method (Patzer 1985) which holds that already a much smaller number of raters produces an accurate measurement of attractiveness. For example, King and Leigh (2009) found that as few as four raters may already lead to a stable attractiveness result. Before the survey, participants were explicitly instructed to base their rating solely on their very first impression and their gut feeling. According to Ballew and Todorov (2007, 17948),

---

[10] Student raters are clearly no representative sample of the voting population in Germany. It is therefore possible that the students rate the appearance of the politicians in a different way than the average voter would do. Due to practical reasons there is nevertheless no good alternative to student raters. For recruiting the students were directly approached during several lectures and asked to participate. We sent every one of these students an email (plus a follow-up mail one week later) including a unique token that allowed a single rater to complete the survey only once. In order to keep the potential bias from student raters low, we used two different student populations from two universities. At the University of Freiburg (Baden-Wuerttemberg) 259 political science students and at the University of Passau (Bavaria) 191 students from communication and cultural studies completed the survey. Albeit the two student populations clearly differed with regard to their academic, geographic and socio-demographic base, the appearance ratings were largely similar (see online appendix O1). This finding suggests that the relative appearance ratings we use in our study are largely stable among raters from different backgrounds. Furthermore, Rosar (2009, 760) rightfully noticed, that even if student raters systematically differ in their ratings compared to the voting population, this would only mean to be a tougher test for the analysis of appearance effects in voting decisions.

"rapid and unreflected face judgements" do a better job at predicting the electoral outcome than if the participants are asked to make a deliberated decision and "think carefully about their choice". Furthermore, Bar et al. (2006) showed "that consistent first impressions can be formed very quickly" (only 39 ms were needed to form a first impression). Our countdown was chosen to support such a quick decision. According to feedback from our pre-testers, ten seconds worked well to allow for comparing the pictures while upholding sufficient pressure to prevent loitering. Additionally, we measured not only who of the two candidates was deemed to be more attractive, competent, and likable, but also – with an accuracy of 10 ms – how long a participant needed to click the respective button. We use these latency times as a measure of ambiguity (see section 3.4).

3.3. Apparent properties of the faces influencing the perception of the three traits

As a first step preceding the main analysis, we test which properties of the candidates' facial pictures are adjuvant for being rated as the more attractive, competent or likable of the two presented candidates.[11] We estimate nine different logistic models: each trait (attractiveness, competence and likability), in each of the three district types (male-vs.-male, female-vs.-female and female-vs.-male races). Figure 2 presents the results. Most of them are fairly expectable: Age has a negative effect on attractiveness in all models, while it has a positive effect on competence, particularly in races with two male candidates. Its effect on likability is much smaller. Glasses tend to make people appear more competent while negatively impacting the other two traits (except for its positive effect on likability in female-vs.-male races. Wearing a beard by trend, reduces the perceived competence of the candidate while it increases the probability to be seen as the more attractive and particularly the more likable person. Wearing a suit or a blazer helps in female-vs.-female races as well as in female-vs.-

---

[11] For this analysis we concentrate on properties of the candidates' faces which are either easily identifiable and measureable or which could at least in theory be changed by the candidate. Thus, we test the effects of gender, wearing glasses or jewelry, but not of facial symmetry.

male races to be perceived as the more competent person, while in male-vs.-male races there is no significant effect. This might of course be due to the small number of male candidates wearing no suit (only 13 out of 364). In female-vs.-male races a short hairstyle for women (virtually all men had short hair) is detrimental to all three traits, while compared to a second female candidate a woman with a short hair style is only perceived as less attractive. For men, a bald head strongly depresses attractiveness, and to a lesser extent also likability but in male-vs.-male races it helps to be perceived as the more competent candidate. The only significant effect of jewelry is probably an unintended one: Wearing jewelry reduces the chances to be rated as the more attractive person in female-vs.-male races. One of the strongest effects can be found for sex of the candidate – in female-vs.-male races men are much less likely to be rated as more attractive or likable than women, but there is no difference in perceived competence.

These findings clearly show that several properties in a picture can shape our perception of a person (sometimes conditioned by gender of the contestants). Furthermore, we tested whether the results presented above differ systematically for male and female raters. This was not the case – there were virtually no differences between the ratings of female and male raters (see figure O2 in the online appendix). It is therefore not necessary to weight the ratings according to the percentage of male/female raters.

Figure 2: Binary Logit: Determinants of Perceived Attractiveness, Competence, and Likability
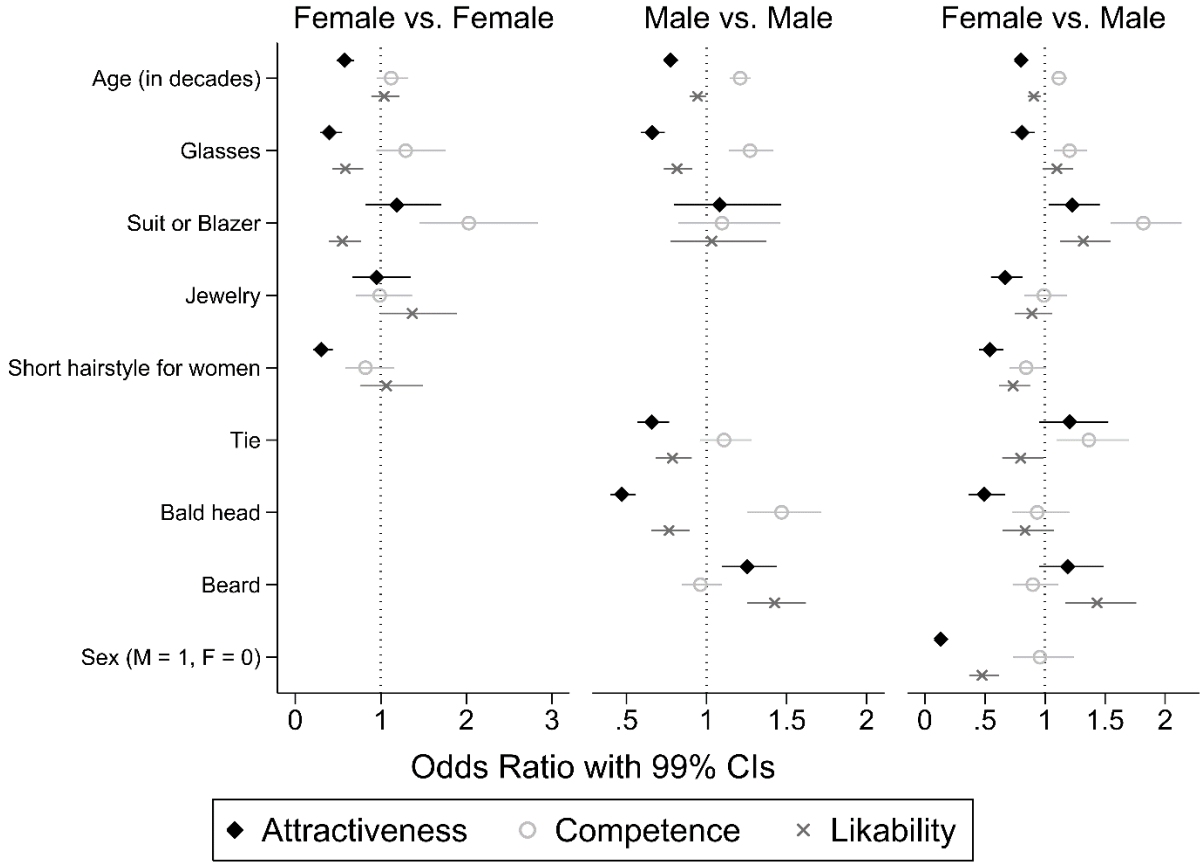
3.4. Ambiguity in the ratings

So far, we have treated all ratings as similar. Each click had the same weight. Yet, it may sometimes be difficult to decide who is more attractive, likable, or competent – i.e. ratings may range from clear-cut to hard-to-decide and ambiguous. This makes the raw ratings difficult to interpret, in particular if a large portion of the raters had trouble to decide. For CATI-research, Bassili argues that "opinions that are expressed quickly are usually more strongly associated with established evaluations" or "are usually more free of conflict than opinions that are expressed slowly" (Bassili 2000, 4). This idea can be adapted to our context: Assuming that raters in doubt require more time to decide, we can use latency times until click as a proxy for ambiguity in the decision which candidate is more attractive (competent, or likable). Furthermore, the most pronounced form of ambiguity can be assumed if a rater

does not click a button at all. In that case he or she could not decide between the candidates, similar to a 50/50-chance to click one of them. Taking these points together, there are three forms of how a high ambiguity should show up in the data: (1) a high number of non-clicks for an electoral district; (2) an equal distribution of clicks across both candidates; (3) a long average time until the button was clicked. For attractiveness, bivariate analysis shows that while all three variables exhibit substantial variation, they are substantially correlated (r = .37 to .56), suggesting that they present a common underlying ambiguity.[12] According to this finding it makes sense to weight the ratings with respect to latency times in order to account for ambiguity. Of course, raters are different regarding their overall clicking-speed. For this reason we centre the latency-times on the rater specific minimal and maximal latency-times for each trait. The fastest click of each rater is assigned a weighting factor of 1.0 while the slowest click gets a weight of .2 (this reflects the assumption that clicking one of the candidates even after a relatively long latency-time does not reflect complete ambiguity on behalf of the rater).[13]

In order to test to what extent the three traits are related in the raters' perception we correlated the weighted scores for the winning candidates of all electoral districts. The results show only a medium strong relation (r =.53) between attractiveness and likability while the other two correlations between competence and attractiveness and competence and likability are below r = .03. The three traits can therefore be seen as relatively independent from each other.

3.5. Dependent variable and controls

---

[12] For the other two traits, likability and competence, the correlations are weaker but still present (see replication file).

[13] The final weighted proportion of clicks for candidate A in district X computes as follows:

$$p_{A,X} = \frac{\sum_{i=1}^{n_X}(r(A)_{i,X} \cdot w_{i,X})}{\sum_{i=1}^{n_X}(w_{i,X})}$$ with: n_X = number of ratings in district X; $w_{i,X}$ = weight for rating i in district X.

$r(A)_{i,X}$ = rating i for candidate A in district X (i.e. 0 or 1);

To analyse the effects of perceived attractiveness, competence, and likability on the performance of direct candidates at the 2013 general election, we estimate OLS regressions. Please note that since our research strategy involves a relative assessment of the three traits (= weighted proportion of clicks for the winning candidate compared to the runner up), this relativity also has to be reflected in the dependent variable. We therefore construct the dependent variable as the difference between the vote share of the winning candidate and the vote share of the runner-up. Taking a look at this variable shows that on average, the winning candidate had a lead of 17 percentage points, while the maximum difference between the winner and the runner-up was 51.4 percentage points.

In addition to the variables that measure perceived appearance, we control for a number of factors characterizing the electoral district: turnout, incumbency status[14] and gender of the candidates[15], share of second votes (i.e. votes for the party), percentage of male citizens, proportion of senior to youth citizens, economic situation (measured via unemployment rate, business tax revenues, and the balance of business registrations and de-registrations) and whether the district was expected to be contested before the 2013 election.[16] Furthermore, we control for candidates holding a doctorate, as studies have shown a positive effect of an academic title in German elections (Schneider and Tepe 2011; Manow and Flemming 2011). Again, all controls regarding the contestants are constructed to reflect the relation between

[14] We use two different operationalizations for incumbency. The first operationalization (incumbency_direct) only counts those canddiates as incumbents who had won the most first votes in 2009 and thus held a „direct mandate" in the legislative term 2009-2013. The second operationalization (incumbency_BT) relaxes this requirement to all candidates who had been a member of the Bundestag between 2009 and 2013 regardless whether they had a „direct mandate" or a „list mandate".

[15] Gender of the electoral race is operationalized using four dummies: male/male; male/female; female/female; female/male (winner/runner up).

[16] This dummy variable was derived from a large inquiry by Spiegel Online (the largest German online news outlet) made for each electoral district as part of their press coverage before the 2013 elections. Districts were marked as contested in light of the 2009 results for first votes and specific case knowledge about issues and events which happened in the district during the interelection period. A map of the marked districts is available at http://www.spiegel.de/politik/deutschland/bundestagswahl-wackel-wahlkreise-2013-a-916641.html

them, not absolute values (e.g. "second votes" is the difference between the vote share of the party of the winning candidate and the party of the runner up). All variables used in the regressions therefore measure at the level of the electoral district. Table O3 in the online appendix gives an overview of all variables, their operationalization and sources.

Following our second research questions, we also test interactions between the three perceived appearance variables and the configuration of the electoral race in terms of gender, incumbency and age.

4. Results

4.1. Main effect models

In a first step we estimate main effects models (see table 1). Models 1 and 3 differ by how we measure incumbency. In model 1 only those candidates who had won the first vote in 2009 are regarded as incumbents, yielding three possibilities: a) winner is incumbent (reference category), b) runner up is incumbent, and c) none of the candidates won the first vote in 2009. In model 3 the incumbency definition is relaxed to cover all candidates that had been members of the Bundestag between 2009 and 2013 (hence four dummies can be created). Model 2 (4) replicates model 1 (3) using a stepwise procedure.

Generally speaking, the explained variance is high and results are consistent across the models.[17] Regarding perceived attractiveness, competence and likability there is a clear finding: Being seen as attractive and, to a lesser extent as competent helps the candidate while likability does not have a significant effect. Since the dependent variable is the difference in the first votes between winner and runner up, the coefficients have to be interpreted as

---

[17] Tests of multicollinearity showed no major problems (max. VIF = 5.1). According to Breusch-Pagan tests, models 1 and 2 exhibit some heteroscedasticity. Therefore, these models are estimated with robust standard errors. Excluding influential observations (Cook's distance > 4/N) does not alter the effects significantly (see table O4 in the online appendix). The results in table 1 can therefore be regarded as robust.

follows: A change in perceived attractiveness from 0 to 1 (i.e. from no one rating the winner as more attractive to 100% rating the winner as the more attractive candidate) increases the distance in the first votes between the winner and the runner up by 2.403 percentage points (model 1). For competence, the effect is slightly weaker and only significant in the stepwise models. The main effect models clearly show that physical appearance in terms of perceived attractiveness (and competence) has significant effects on the vote shares for direct candidates even when controlling for other relevant factors such as incumbency or the percentage of second votes for the winning candidate's party (which is by far the strongest predictor when interpreting the betas see table O5 in the online appendix). Likability only becomes significant when attractiveness is omitted.

Gender plays a minor role: only in female-vs.-male races, the difference between the winner and the runner up (regardless whether the woman or the man wins) is slightly larger than in the reference category of female-vs.-female districts. The other controls show that in electoral districts with a high share of senior-to-youth citizens, the difference between the winner and the runner up decreases. The same applies to those districts that had a priori been classified as contested. Interestingly, the difference in age of the candidates is only slightly significant in models 1 and 2. By trend, the older the winning candidate is compared to the runner up, the smaller is her lead. The following interaction models investigate this result further.

Table 1: Main effects models

	Model 1	Model 2 (stepwise)	Model 3	Model 4 (stepwise)
Perceived physical appearance				
Attractiveness	2.403***	2.185***	2.711***	3.351***
	(0.795)	(0.612)	(0.904)	(0.696)
Competence	1.475	2.061**	2.156*	
	(1.098)	(0.949)	(1.253)	
Likability	0.883		0.872	
	(0.928)		(1.087)	
Incumbency dummies				
Incumbency_BT (only runner up)	-4.688***	-4.442***		
	(0.818)	(0.785)		
Incumbency_BT (both cand.)	-2.531***	-2.576***		
	(0.488)	(0.479)		
Incumbency_BT (none of the cand.)	-2.275***	-2.067***		
	(0.502)	(0.479)		
Incumbency_direct (runner up)			-3.832***	-3.773***
			(0.798)	(0.778)
Incumbency_direct (none of the cand.)			-1.908***	-1.699***
			(0.502)	(0.467)
Gender dummies				
Winner: male & Runner up: male	1.139		1.189	
	(0.860)		(0.820)	
Winner: male & Runner up: female	1.589		1.785*	0.873**
	(0.966)		(0.907)	(0.438)
Winner: female & Runner up: male	1.349		1.665*	
	(0.927)		(0.916)	
Controls	1.139		1.189	
Age difference	-0.0324*	-0.0350**	-0.0199	
	(0.0173)	(0.0161)	(0.0183)	
Doctorate	0.275		0.0864	
	(0.330)		(0.399)	
Contested electoral district	-2.495***	-2.552***	-2.455***	-2.303***
	(0.589)	(0.546)	(0.514)	(0.493)
Second votes 2013	0.975***	0.981***	0.973***	0.981***
	(0.0246)	(0.0198)	(0.0236)	(0.0179)
Turnout	-0.103*		-0.110	
	(0.0587)		(0.0678)	
Business tax revenues (in 1000 Euro per capita)	0.276		0.255	
	(0.697)		(0.855)	
Business registrations/deregistrations (per 1000 persons)	-0.0515		-0.0222	
	(0.184)		(0.190)	
Unemployment rate	-0.0718		-0.0742	
	(0.0995)		(0.108)	
Percentage male population	-0.528		-0.539	
	(0.339)		(0.348)	
Senior-to-youth-rate	-1.910	-1.608*	-2.555**	-2.005**
	(1.241)	(0.858)	(1.271)	(0.896)
Constant	35.99*	3.273***	36.43*	3.167***
	(19.59)	(1.158)	(20.08)	(1.185)
Observations	259	259	259	259
R-squared	0.947	0.945	0.941	0.938

Standard errors in parentheses (Models 1 and 1 with robust standard errors due to heteroscedasticity)
*** p<0.01, ** p<0.05, * p<0.1

4.2. Interaction models

All interaction models are based on the main effects model 1. The interactions were tested in separate models. Table 2 gives an overview of all significant interactions. The complete models can be found in table O6 in the online appendix. For interactions with the metric variable "age difference", significance is assumed if there is a marginal effect significantly different from 0 at some point within the range between the empirical minimum and maximum of age difference (-30 to 30). Figures 3 and 4 present the two significant interactions with the incumbency dummy and the one between gender and likability using predicted value plots. The interactions with "age difference" are presented as marginal effect plots in figure 5.

Table 2: Overview of all significant interactions

	Attractiveness	Competence	Likability
Incumbency_BT			
Winner	0	0	0
Runner up	+	0	0
No one	0	0	0
Both	0	0	+
Gender			
Female vs. female	0	0	+
Male vs. male	0	0	0
Female vs. male	0	0	0
Male vs. female	0	0	0
Age difference	-	0	0

+: positive interaction effect, -: negative interaction effect, 0: no significant interaction

The interaction between incumbency and attractiveness shows the expected effect: Only when the 2013 runner up had been a member of the Bundestag in 2009, attractiveness plays a major role. This means that attractiveness is more important for candidates who were new to the contest in 2013 and did not have the incumbency advantage. The three other incumbency dummies "only winner is incumbent", "no one of the candidates is incumbent" and "both

candidates had been members of the Bundestag" show no significant interaction effect.[18] While for competence there is no significant interaction with incumbency status, figure 3b shows that likability becomes relevant in cases where both candidates had been members of the Bundestag before.[19]

Figure 3: Interactions between Incumbency_BT and Attractiveness (a)/Likability (b) (Predicted Values)

a)
Predicted value
(= difference in percentage points between the winner and the runner up)

— winner is incumbent_BT
- - runner up is incumbent_BT
····· 95% CI

Attractiveness
(percentage of raters regarding the winner as beeing more attractive than the runner up)

b)
Predicted value
(= difference in percentage points between the winner and the runner up)

— winner is incumbent_BT
- - both are incumbents_BT
····· 95% CI

Likability
(percentage of raters regarding the winner as beeing more likable than the runner up)

The third significant interaction effect is between gender and likability. Likability does not show any effect in districts where two male candidates compete or where a woman competes against a man (regardless whether the woman is winner or runner up), but when female candidates compete, likability becomes a relevant factor.[20] At this point, we can only speculate why likability seems to work as a separate dimension for women, but not for men. Yet, the expectation that male and female candidates would generally profit to a different

---

[18] In order to increase readability the interactions for "no one" and "both candidates are incumbents" have been omitted from the figure. This interaction line is not presented in figure 5 to increase the readability of the plot. It has nearly the same position as the line for "winner is incumbent".

[19] Again the two other non significant interactions („runner up is incumbent" and „no one is incumbent") have been omitted in order to increase readability

[20] Again figure 6 only presents the interaction line for male vs. male districts. The lines for female-vs.-male and male-vs.-female districts differ only marginally from that line.

extent from their physical appearance has to be denied. Neither the strong effect of attractiveness nor the weaker one for perceived competence is conditioned by the candidates gender.

Figure 4: Interaction between Gender and Likability (Predicted Values)
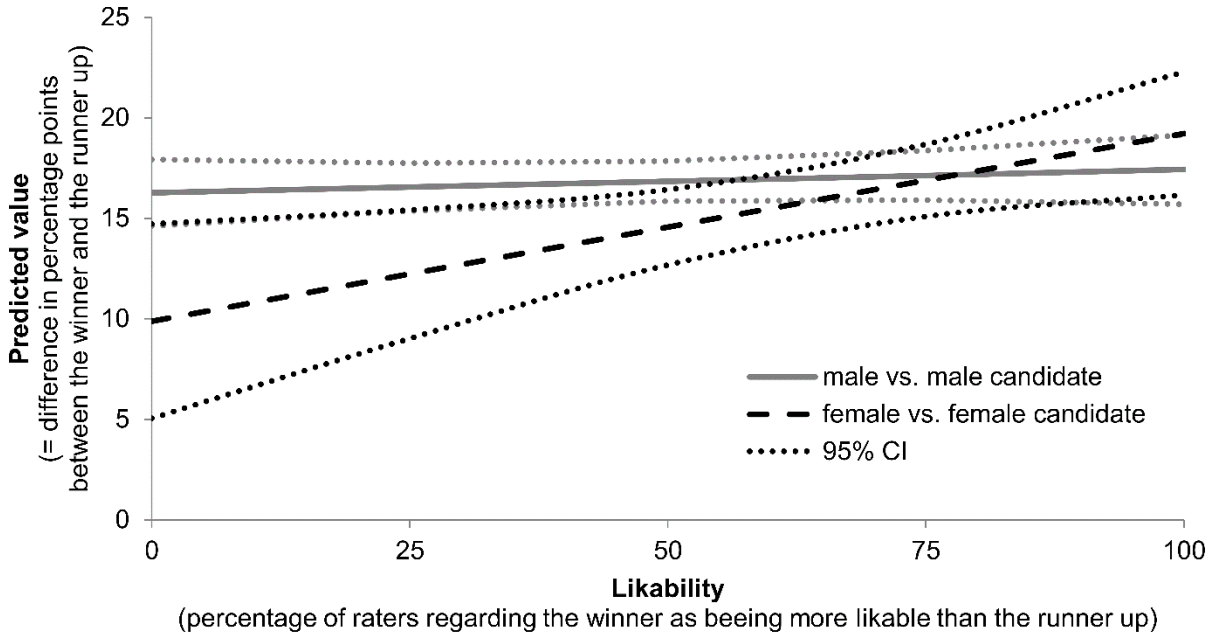

The last interaction tested is between the age difference of winner and runner up and the three traits. The marginal effect plots indicate that there are significant interactions when the winning candidate is younger than his or her contestant. We find the strongest effect between age and attractiveness. This means that for comparatively young candidates attractiveness plays a larger role than for their older opponents. For a winning candidate who is 20 years younger than the runner up an increase in the attractiveness score from 0 to 1 would mean a rise in the difference between the first vote shares of the two by about 3.5 percentage points. If both candidates have the same age, the attractiveness coefficient drops to 2.0, and if the runner up is 10 years younger than the winner, there is no significant attractiveness advantage detectable any longer. For competence as well as for likability the interactions show the same trend, yet they fall short of significance.

Figure 5: Interactions between Age Difference and (a) Attractiveness, (b) Competence, and (c) Likability
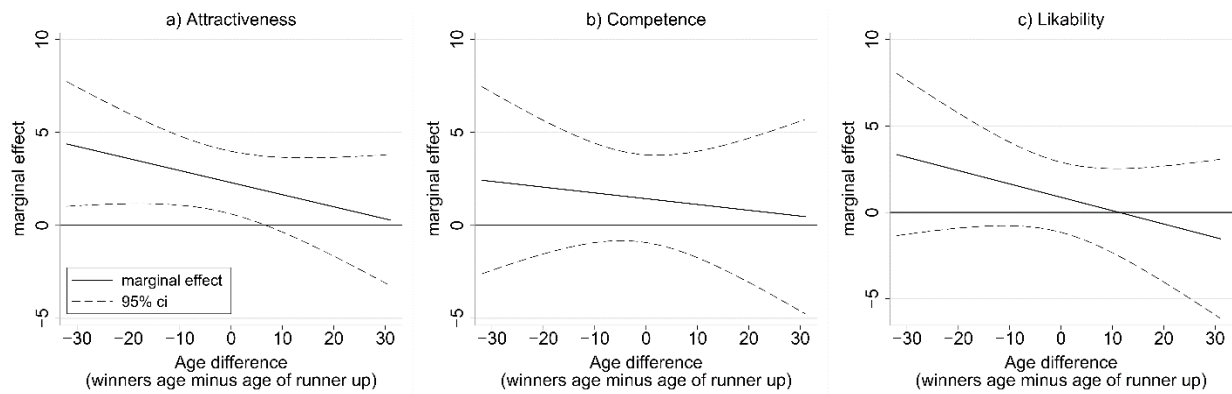
5. Summary and discussion

5.1 Summary of the results

We began this article with the two questions whether perceived attractiveness, competence or likability of direct candidates to the German Bundestag can serve as relevant predictors for their electoral success (which of these traits matters most), and whether potential appearance effects may be conditioned by incumbency status, gender and age of the candidates.

To answer these questions, we conducted an online survey among 449 students from different academic and geographic backgrounds. They indicated for all the 259 electoral districts where no well-known candidate competed, whether in their view the winner of the first vote in the 2013 general elections appeared as being more attractive, more competent and more likable than the runner up. The raters did not know that they were rating politicians. Using latency times as an indicator of ambiguity, we aggregated the ratings at district level to form a measure of the perceived attractiveness, competence and likability of the winning candidate relative to the second placed. This relative measurement is more faithful to the situation in a district (voters will compare available candidates instead of judging them in the way a jury in

a beauty contest would) and distinguishes our article from other approaches which measure perceived beauty (or other traits) in an absolute way.

Prior to the main analysis we tested which apparent features of the candidates' pictures are positively correlated with being perceived as more attractive, competent or likable. Most of the effects we find are very plausible: Young persons have better odds to be rated as being more attractive, while wearing glasses boosts competence and having a beard makes the candidate likable, to name but a few. Yet, these findings would be quite irrelevant from a political science point of view if the three appearance based traits had no impact on the election results.

Using the difference between the first vote shares of the winner and the second placed candidate as dependent variable, we find that attractiveness, and to a lesser extent competence, both positively affect the vote share. Likability, in contrast, shows no significant effect. These results are robust to controlling for potentially relevant factors such as different kinds of incumbency, age and gender of the candidates, turnout, economic situation within the electoral district and even the share of second votes which has by far the largest effect. In total, our model explains more than 94% of the variance.

Relating our findings back to the literature, our results neither fully support the idea of Todorov and colleagues that competence wins out nor the idea of Berggren and colleagues that attractiveness is the only relevant factor. Rather, both attributes seem to work independently of each other and their effects add up – yet, the effect of perceived attractiveness is stronger and always significant, whereas the competence effect is weaker and falls short of significance in some models. Our study thus lends additional support to the idea of a beauty premium, even under tough conditions, but also shows that candidates are not just evaluated on the basis of attractiveness alone. Rather competence as a factor more closely related to their role as politicians also plays a part, even if it may merely be inferred from a

photograph. According to our findings, when parties want to toss up a contested district, they should send a candidate that is both good and competent looking.

In the second part of the main analysis we tested whether the appearance effects are conditioned by gender, incumbency and age as suggested by earlier studies. Indeed, we find some significant interactions. Attractiveness plays a stronger role for those candidates who do not have the incumbency advantage. This makes sense insofar, as voters can rely on other information when making up their mind about incumbents (they probably know more about them), but they have to hark back to other possibilities for judging challengers, and attractiveness seems to work well in that case. This result is in line with earlier work (King and Leigh 2009). Yet, our models do not corroborate results from Praino and colleagues (2014) who assume that competence is more important for inter-gender races while in intra-gender races it is attractiveness. Both traits show consistently positive effects on the first vote share, regardless of the gender of the candidates. We find the only significant interaction between one of the appearance variables and gender between all-female districts and likability. This means that while in general likability shows no significant effect, at least as long as attractiveness is controlled for, looking likable helps a female candidate when competing against another woman. The final interactions that we tested show that young candidates can benefit more from a high attractiveness compared to their opponents than older candidates.

5.2 Practical and normative implications

We can conclude that for elections to the German Bundestag, with its comparatively low personalization and at the same time relatively high party identification which in principle should suppress effects of candidates' physical appearance, beauty nevertheless "pays" and is never "beastly". To a somewhat lesser extent, the same is true for perceived competence. Particularly for challengers and young politicians their physical appearance is a factor they

themselves as well as the parties nominating them can hardly neglect in the tough fight for the first vote. Our study nevertheless also shows that while appearance matters, it is not everything – other factors such as incumbency status or the share of second votes influence electoral success to a bigger extent – yet, particularly in tight races, visual appearance can make the difference. For candidates themselves it certainly matters that their facial attributes may help or hinder being perceived as attractive, competent or likable and that these perceptions have an impact on their chances of success at the ballot box. Candidates aware of the mechanism could actively try to shape their appearance with an eye towards polls, and, ultimately, electoral results (e.g. if their competence rating is low they could start to wear glasses, or they could grow a beard to become more attractive). As long as the consequences of the effects we found are such simple modifications of the appearances and looks of a candidate, this probably does not present a major problem from the point of democratic theory. There really is no argument against presenting oneself in way that appeals to others – virtually everybody does so to some extent. Yet, the picture changes if parties used this information for establishing a candidate selection process that puts special emphasis on physical appearance to just gain that extra momentum at the ballot. Once actual qualification or ideological positions of the candidates run the risk of ending up in the backseat compared to outward appearance, we can expect to see effects on both the quality of the campaigns and the resulting selection of political elites. Campaigns that are all tinsel and glitter with candidates only selected because of their appearance will almost certainly take their toll on the confidence into the political system and its representatives as a central element of the democratic process. For Germany, it seems that at the moment this is not the case, yet mediatisation and personalization of politics well known from the United States are also on their way here and could lead to the impression of an increasing superficiality in the political arena. For future studies it could therefore be interesting to test whether there has been an increase in the effects of candidates' appearances, and, whether age-, period-, or cohort effects

play a role too. Another interesting question to focus on, based on the results from this study –
which only tackled elections to the German Bundestag – could be to systematically compare
appearance effects between different political systems and political cultures. For example: Is
there a universal effect of attractiveness and competence on the chances to get elected, even if
in different cultures different facial characteristics may shape the perception of being
attractive or competent? While this study has shown the importance of dealing with physical
appearance in election studies, these and other related questions surely deserve a more
thorough inspection.

References

Armstrong, J. Scott, Green, Kesten C., Jones, Randall J., Jr., Wright, Malcolm J., 2010. Predicting elections from politicians' faces. *International Journal of Public Opinion Research 22*, 511-522.

Arzheimer, Kai, 2006. Dead men walking? Party identification in Germany, 1977-2002. *Electoral Studies 25(4)*, 791-807.

Atkinson, Matthew D., Enos, Ryan D., Hill, Seth J., 2009. Candidate faces and election outcomes: is the face-vote correlation caused by candidate selection? *Quarterly Journal of Political Science 4(3)*, 229-249.

Ballew, Charles C., Todorov, Alexander, 2007. Predicting political elections from rapid and unreflected face judgements. *Proceedings of the National Academy of Sciences 104(46)*, 17948-17953.

Barrett, Andrew W., Barrington, Lowell W., 2005. Is a picture worth a thousand words? Newspaper photographs and voter evaluations of political candidates. *The Harvard International Journal of Press/Politics 10(4)*, 98-113.

Bassili, John N., 2000. Reflections on Response Latency Measurement in Telephone Surveys, *Political Psychology 21(1)*, 1-6.

Berggren, Niclas, Jordahl, Henrik, Poutvaara, Panu, 2010. The looks of a winner: Beauty and electoral success. *Journal of Public Economics 94(1-2)*, 8-15.

Berglund, Frode, Holmberg, Sören, Schmitt, Hermann, Thomassen, Jacques, 2005. Party Identification and Party Choice. In: Thomassen, Jacques: The European Voter, Oxford University Press, Oxford, pp. 105-123.

Bishin, Benjamin G., Stevens, Daniel, Wilson, Christian, 2006. Character counts? Honesty and fairness in election 2000. *Public Opinion Quarterly 70(2)*, 235-248.

Bóo, Florencia López., Rossi, Martín A., Urzúa, Sergio S., 2013. The labor market return to an attractive face: Evidence from a field experiment, *Economics Letters 118*, 170-172.

Brettschneider, Frank, 2002. Spitzenkandidaten und Wahlerfolg. Personalisierung, Kompetenz, Parteien. Ein internationaler Vergleich. Westdeutscher Verlag, Wiesbaden.

Bull, Ray, Hawkes, Caroline, 1982. Judging Politicians by their Faces. *Political Studies 30(1)*, 95-101.

Campbell, Angus, Converse, Philip E., Miller, Warren E., Stokes, Donald E., 1960: The American voter. Wiley, New York.

Dalton, Russell J., 2000. The decline of party identifications. In: Dalton, Russell J., Wattenberg, Martin P. (Eds.), Parties without partisans. Oxford University Press, Oxford, pp. 19-36.

Dancey, Logan, Sheagley, Geoffrey, 2013. Heuristics Behaving Badly: Party Cues and Voter Knowledge. *American Journal of Political Science 57(2)*, 312-325.

Dion, Karen, Berscheid, Ellen, Walster, Elaine, 1972. What is beautiful is good. *Journal of Personality and Social Psychology 24(3)*, 285-290.

Dukake, Imke, Kiechle, Thomas, Klein, Markus, Rosar, Ulrich, 2012. Schöne Schüler. Schöne Noten? Eine empirische Untersuchung zum Einfluss der physischen Attraktivität von Schülern auf die Notenvergabe durch das Lehrpersonal, *Zeitschrift für Soziologie 41*, 142-161.

Efran, Michael G., Patterson E.W.J., 1974. Voters vote beautiful: the effect of physical appearance on a national election. *Canadian Journal of Behavioural Science 6(4)*, 352-356.

Gordon, Rachel A., Crosnoe, Robert, Wang, Xue, 2014. Physical attractiveness and the accumulation of social and human capital in adolescence and young adulthood: Assets and distractions. Wiley, New York.

Hamermesh, Daniel S., Biddle, Jeff E., 1994. Beauty and the labor-market. *The American Economic Review 84*, 1174-1194.

Heilman, Madeline E., Saruwatari, Lois 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.

Kaase, Max, 1994. Is there personalization in politics? Candidates and voting behavior in Germany. *International Political Science Review 15(3)*, 211-230.

Kenrick, Douglas T., Gutierres, Sara E., 1980. Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology 38(1)*, 131-140.

King, Amy, Leigh, Andrew, 2009. Beautiful politicians. *KYKLOS 62(4)*, 579-593.

Klein, Markus, Rosar, Ulrich, 2005. Physische Attraktivität und Wahlerfolg. Eine empirische Analyse am Beispiel der Wahlkreiskandidaten bei der Bundestagswahl 2002. *Politische Vierteljahresschrift 46(2)*, 263-287.

Johnson, Stefanie K./, Podratz, Kenneth E., Dipboye, Robert L., Gibbons, Ellie, 2010. Physical attractiveness biases in ratings of employment suitability: Tracking down the „Beauty is Beastly" effect. *Journal of Social Psychology 150(3)*, 301-318.

Langlois, Judith H., Ritter, Jean M., Casey, Rita J., Sawin, Douglas B., 1995. Infant attractiveness predicts maternal behaviors and attitudes. *Developmental Psychology 31(3)*, 464-472.

Lawson, Chapell, Lenz, Gabriel S., Baker, Andy, Myers, Michael, 2010. Looking like a winner: Candidate appearance and electoral success in new democracies. *World Politics 62(4)*, 561-593.

Little, Anthony C., Burriss, Robert P., Jones, Benedict C., Roberts, S. Craig, 2007. Facial appearance affects voting decisions. *Evolution of Human Behavior 28*, 18-27.

Manow, Philip, Flemming, Peter, 2011. Der Titel als politisches Distinktionsmerkmal? Eine Untersuchung akademischer und adeliger Wahlbewerber zum Deutschen Bundestag 1949-2009. *Zeitschrift für Politikwissenschaft 21(4)*, 531-551.

McDermott, Monika L., 2009. Voting for myself. Candidate and voter group associations over time. *Electoral Studies 28(4)*, 606-614.

McLellan, Barbara, McKelvie Stuart J., 1993. Effects of Age and Gender on Perceived Facial Attractiveness. *Canadian Journal of Behavioural Science 25(1)*, 135-142.

Ohr, Dieter, Klein, Markus, Rosar, Ulrich, 2013. Bewertungen der Kanzlerkandidaten und Wahlentscheidung bei der Bundestagswahl 2009. In: Weßels, Bernhard, Schoen, Harald, Gabriel, Oskar W. (eds.), *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2009*. Springer Fachmedien, Wiesbaden, 206-230.

Olivola, Christopher Y./Todorov, Alexander, 2010. Elected in 100 milliseconds: Appearance-Based Trait Inferences and Voting. *Journal of Nonverbal Behavior 34*, 83-110.

Pappi, Franz Urban, Shikano, Susumu, 2001. Personalisierung der Politik in Mehrparteiensystemen am Beispiel deutscher Bundestagswahlen seit 1980. *Politische Vierteljahresschrift 42(3)*, 355-387.

Praino, Rodrigo, Stockemer, Daniel, Ratis, James, 2014. Looking Good or Looking Competent? Physical Appearance and Electoral Success in the 2008 Congressional Elections. *American Politics Research 42(6)*, 1096-1117.

Rosar, Ulrich, Klein, Markus, Beckers, Tilo, 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, 64-79.

Rosar, Ulrich, Hagenah, Jörg, Klein, Markus, 2010. Physische Attraktivität und individuelles Leistungsverhalten. *Soziale Welt 61(1)*, 51-68.

Rule, Nicholas O., Ambady, Nalini, Adams, Reginald B., Ozono, Hiroko , Nakashima, Satoshi, Yoshikawa, Sakiko, Watabe, Motoki, 2010. Polling the face: Predictionand consensus across cultures. *Journal of Personality and Social Psychology 98(1)*, 1-15.

Schneider, Sebastian, Tepe, Markus, 2011. Dr. Right and Dr. Wrong. Zum Einfluss des Doktortitels auf den Wahlerfolg von Direktkandidaten bei der Bundestagswahl 2009. *Politische Vierteljahresschrift 52*, 248-285.

Spezio, Michael L., Loesch, Laura, Gosselin, Frédéric, Mattes, Kyle, Alvarez, Michael R., 2012. Thin-slice decisions do not need faces to be predictive of election outcomes. *Political Psychology 33(3),* 331-341.

Todorov, Alexander, Mandisodza, Anesu N., Goren, Amir, Hall, Crystal C., 2005. Inferences of competence from faces predict election outcomes. *Science 308*, 1623-1626.

Beauty contest revisited – the effects of perceived attractiveness, competence and likability on the electoral success of German MPs

Online Appendix

O1: Comparison between the appearance ratings of the Freiburg and Passau student raters

District type

Odds Ratio with 99% CIs
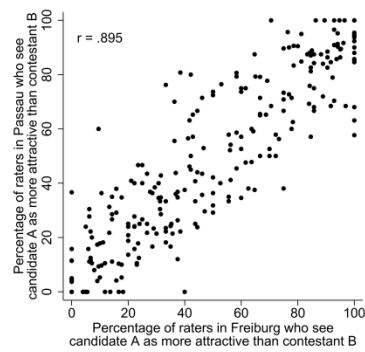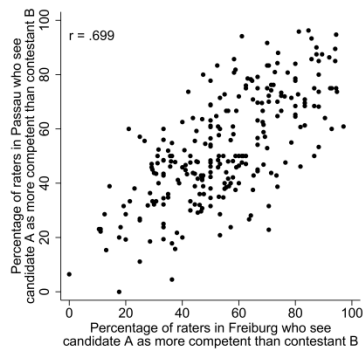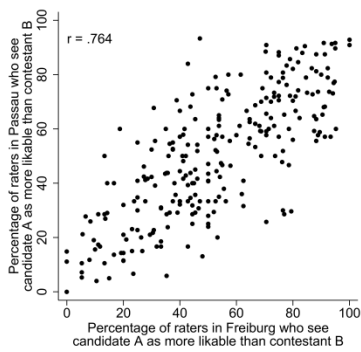
○ Female rater   ● Male rater

O3: Variable overview
(all variables measure on the level of the electoral district)

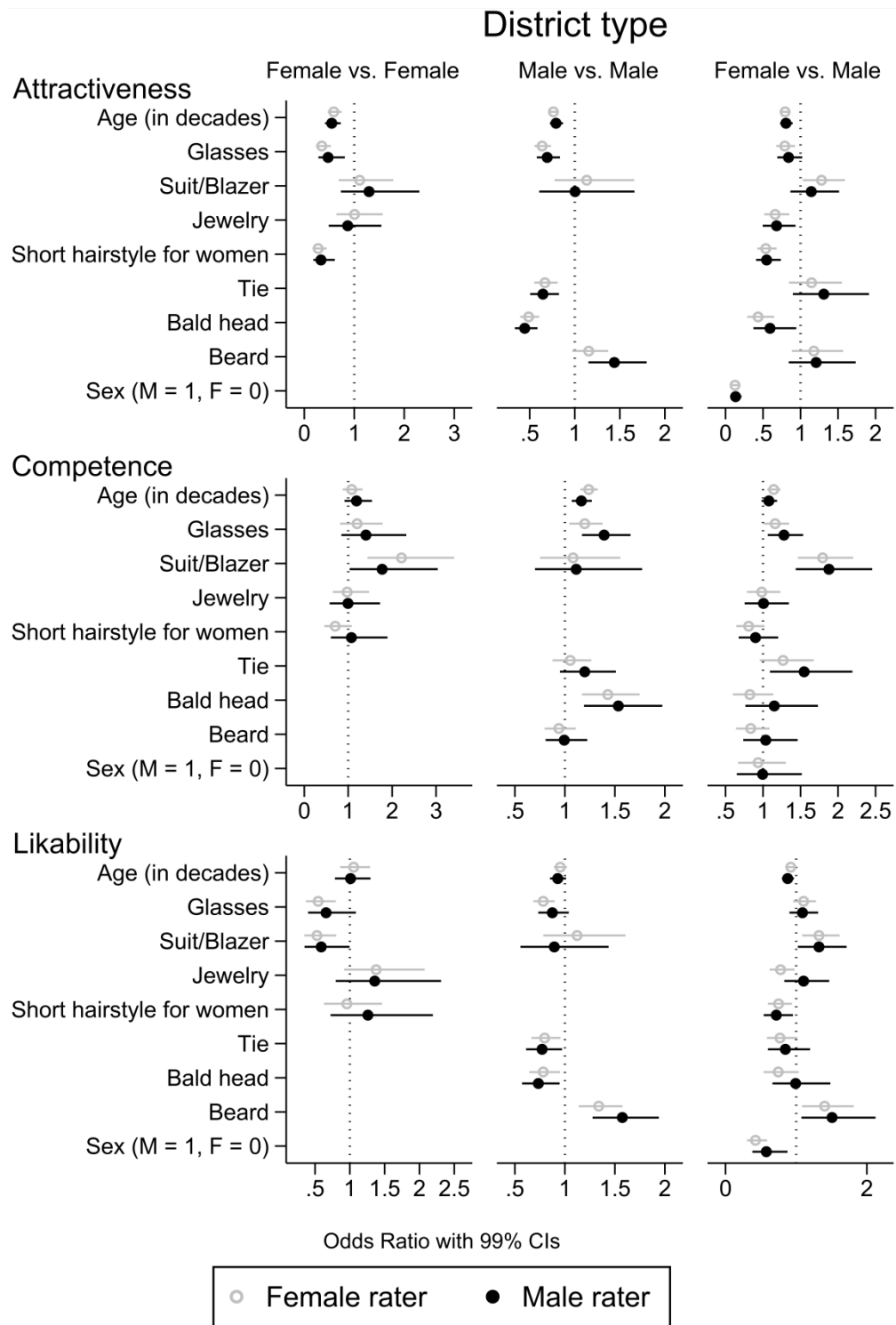| Variable	operationalization	min / max	mean (sd)	source
		For dummies: % of cases		
Dependent Variable	Vote share winner – vote share runner up	0 / 51.4	17.577 (11.986)	Electoral management body Germany (Bundeswahlleiter)
Attractiveness	Latency weighted mean of clicks for the winning candidate compared to the runner-up	0 / 1	0.500 (0.295)	Own online survey
Competence		0 / 0.952	0.532 (0.182)	
Likability		0 / 0.956	0.509 (0.210)	
Incumbency_direct	Three dummies indicating who of the candidates (if any) had won the direct constituency seat	Incumbent _direct = … Winner 2013: 70.27% Runner up 2013: 7.34% Neither winner nor runner up: 22.39%		Electoral management body Germany (Bundeswahlleiter)
Incumbency_BT	Four dummies indicating who of the candidates (if any) had been members of the German Bundestag in the legislative term 2009-2013	Incumbent_BT = … Winner 2013: 49.81% Runner up 2013: 10.04% Both Winner and Runner up: 23.17% Neither winner nor runner up: 16.99%		Electoral management body Germany (Bundeswahlleiter)
Gender	Four dummies indicating the gender configuration of the race	Winner / Runner up: ♂ / ♂ 47.49% ♂ / ♀ 32.82% ♀ / ♂ 12.74% ♀ / ♀ 6.95%		Electoral management body Germany (Bundeswahlleiter)
Age difference	Age of the winner minus age of the runner up	-32 / 31	-0.398 (13.089)	Candidates CVs
Doctorate	1: Only winner has a doctor's degree 0: Both candidates or none of them have a doctor's degree -1: Only runner up has a doctor's degree	-1 / 1	0.046 (0.496)	Candidates CVs
Contested electoral district	Dummy identifying districts that had been regarded as contested before the 2013 election	21.24%		http://www.spiegel.de/politik/deutschland/bundestagswahl-wackel-wahlkreise-2013-a-916641.html (last 05.05.2015)
Second votes 2013	Second votes for party of the winner minus second votes for the party of the runner up (in %)	-10.353 / 43.580	16.159 (11.214)	Electoral management body Germany (Bundeswahlleiter)
Turnout	In %	58.9 / 79.8	71.378 (3.898)	Electoral management body Germany (Bundeswahlleiter)
Business tax revenues	In 1000 Euro per capita	0.149 / 1.803	0.479 (0.274)	Federal Statistical Office Germany
Business registrations/ deregistrations	Per 1000 persons	-3.3 / 4.7	1.195 (1.243)	Federal Statistical Office Germany
Unemployment rate	In %	2.0 / 14.0	6.763 (2.935)	Federal Statistical Office Germany
Percentage male population	In %	46.644 / 51.657	49.179 (0.649)	Federal Statistical Office Germany
Senior-to-youth-rate	$\dfrac{\text{Share of seniors } (> 60)}{\text{Share of youth } (<25)}$	0.657 / 1.861	1.128 (0.220)	Federal Statistical Office Germany

O4: Main models with influential cases excluded

	Model 1	Model 2 (stepwise)	Model 3	Model 4 (stepwise)
Perceived physical appearance				
Attractiveness	2.987***	3.127***	2.755***	2.820***
	(0.710)	(0.580)	(0.710)	(0.570)
Competence	0.566		0.661	
	(0.989)		(0.993)	
Likability	0.602		0.709	
	(0.846)		(0.845)	
Incumbency dummies				
Incumbency_BT (only runner up)			-4.220***	-3.389***
			(0.597)	(0.523)
Incumbency_BT (both cand.)			-2.529***	-2.010***
			(0.389)	(0.404)
Incumbency_BT (none of the cand.)			-2.398***	-2.198***
			(0.441)	(0.402)
Incumbency_direct (runner up)	-4.250***	-3.939***		
	(0.660)	(0.729)		
Incumbency_direct (none of the cand.)	-1.704***	-1.702***		
	(0.387)	(0.381)		
Gender dummies				
Winner: male & Runner up: male	2.189***	1.271*	2.347***	
	(0.668)	(0.656)	(0.675)	
Winner: male & Runner up: female	2.808***	2.154***	2.664***	0.876***
	(0.739)	(0.713)	(0.742)	(0.336)
Winner: female & Runner up: male	3.621***	2.350***	2.811***	0.991**
	(0.743)	(0.743)	(0.741)	(0.485)
Controls				
Age difference	-0.00607		-0.0263*	
	(0.0146)		(0.0146)	
Doctorate	0.412		0.451	0.553*
	(0.309)		(0.309)	(0.323)
Contested electoral district	-2.361***	-2.287***	-2.739***	-2.430***
	(0.420)	(0.420)	(0.408)	(0.433)
Second votes 2013	1.020***	1.007***	1.010***	1.005***
	(0.0197)	(0.0149)	(0.0194)	(0.0159)
Turnout	-0.0744		-0.0832	
	(0.0538)		(0.0529)	
Business tax revenues	-0.423		0.0505	
(in 1000 Euro per capita)	(0.690)		(0.657)	
Business registrations/deregistrations	-0.00297		-0.0195	
(per 1000 persons)	(0.151)		(0.151)	
Unemployment rate	0.0772		0.0941	
	(0.0852)		(0.0862)	
Percentage male population	-0.436		-0.179	
	(0.280)		(0.280)	
Senior-to-youth-rate	-2.686***	-1.558**	-1.828*	-1.334**
	(1.018)	(0.740)	(1.015)	(0.653)
Constant	27.27*	1.153	14.94	2.755***
	(16.18)	(1.230)	(15.98)	(0.941)
Observations	238	241	242	242
R-squared	0.966	0.962	0.966	0.964

To test whether the most influential cases alter the main effects, all cases with a Cook's distance greater than 4/N were excluded from the models (see Stata 2013: 1881). Standard errors in parentheses (Models 1 and 2 with robust standard errors due to heteroscedasticity) *** p<0.01, ** p<0.05, * p<0.1

O5: Main models with betas

	Model 1 Beta	Model 2 Beta	Model 3 Beta	Model 4 Beta
Perceived physical appearance				
Attractiveness	0.0592***	0.0538***	0.0668***	0.0826***
Competence	0.0224	0.0313**	0.0327*	
Likability	0.0155		0.0153	
Incumbency Dummies				
Incumbency_BT (only runner up)	-0.118***	-0.112***		
Incumbency_BT (both cand.)	-0.0893***	-0.0908***		
Incumbency_BT (none of the cand.)	-0.0714***	-0.0649***		
Incumbency_direct (runner up)			-0.0835***	-0.0822***
Incumbency_direct (none of the cand.)			-0.0665***	-0.0592***
Gender dummies				
Winner: male & runner up: male	0.0475		0.0496	
Winner: male & runner up: female	0.0624*		0.0700*	0.0343*
Winner: female & runner up: male	0.0376		0.0464	
Controls				
Age difference	-0.0354*	-0.0382**	-0.0217	
Doctorate	0.0114		0.00357	
Contested electoral district	-0.0853***	-0.0872***	-0.0839***	-0.0787***
Second votes 2013	0.913***	0.918***	0.910***	0.918***
Turnout	-0.0337		-0.0359	
Business tax revenues (in 1000 Euro per capita)	0.00632		0.00583	
Business registrations/deregistrations (per 1000 persons)	-0.00534		-0.00230	
Unemployment rate	-0.0176		-0.0182	
Percentage male population	-0.0286		-0.0292	
Senior-to-youth-rate	-0.0351	-0.0296*	-0.0470**	-0.0369**
Observations	259	259	259	259
R-squared	0.947	0.945	0.941	0.938

Normalized beta coefficients
*** p<0.01, ** p<0.05, * p<0.1

O6: Interaction Models

	Int. Model 1 Incumbency_BT x attractiveness	Int. Model 2 Incumbency_BT x competence	Int. Model 3 Incumbency_BT x likability	Int. Model 4 Gender x attractiveness	Int. Model 5 Gender x competence	Int. Model 6 Gender x likability	Int. Model 7 Age x attractiveness	Int. Model 8 Age x competence	Int. Model 9 Age x likability
Perceived physical appearance									
Attractiveness	1.425	2.454***	2.422***	2.559**	2.371***	2.485***	2.301***	2.427***	2.424***
	(1.042)	(0.865)	(0.864)	(0.996)	(0.865)	(0.865)	(0.861)	(0.862)	(0.859)
Competence	1.181	0.961	1.344	1.525	-0.0197	1.699	1.624	1.436	1.588
	(1.205)	(1.495)	(1.201)	(1.225)	(1.529)	(1.204)	(1.202)	(1.204)	(1.202)
Likability	0.986	0.818	-0.241	0.875	0.760	1.156	0.960	0.925	0.873
	(1.040)	(1.042)	(1.296)	(1.046)	(1.039)	(1.385)	(1.035)	(1.042)	(1.035)
Incumbency dummies									
Incumbency_BT (runner up)	-7.001***	-4.326**	-7.640***	-4.703***	-4.598***	-4.704***	-4.659***	-4.665***	-4.713***
	(1.338)	(1.957)	(1.994)	(0.674)	(0.672)	(0.665)	(0.668)	(0.672)	(0.668)
Incumbency_BT (none of the cand.)	-2.660**	-3.953**	-1.825	-2.278***	-2.317***	-2.091***	-2.272***	-2.263***	-2.279***
	(1.130)	(1.806)	(1.385)	(0.545)	(0.549)	(0.545)	(0.540)	(0.543)	(0.541)
Incumbency_BT (both cand.)	-3.546***	-3.073**	-4.889***	-2.538***	-2.560***	-2.387***	-2.522***	-2.541***	-2.502***
	(0.884)	(1.424)	(1.231)	(0.478)	(0.475)	(0.475)	(0.474)	(0.476)	(0.475)
Gender dummies									
Winner: female & Runner up: female	-1.218	-1.201	-1.428*	-0.929	-1.505	-6.379**	-1.135	-1.152	-1.181
	(0.780)	(0.787)	(0.782)	(2.448)	(2.190)	(2.540)	(0.778)	(0.782)	(0.780)
Winner: male & Runner up: female	0.500	0.415	0.394	0.705	-1.440	1.187	0.458	0.446	0.431
	(0.449)	(0.451)	(0.450)	(0.790)	(1.482)	(1.062)	(0.447)	(0.449)	(0.448)
Winner: female & Runner up: male	0.237	0.0620	0.179	0.183	-1.763	1.971	0.291	0.185	0.220
	(0.596)	(0.615)	(0.597)	(1.684)	(1.610)	(1.760)	(0.599)	(0.601)	(0.596)
Controls									
Age difference	-0.0310*	-0.0334*	-0.0323*	-0.0333*	-0.0314*	-0.0319*	0.000623	-0.0154	0.00916
	(0.0175)	(0.0178)	(0.0176)	(0.0178)	(0.0176)	(0.0176)	(0.0303)	(0.0440)	(0.0404)
Doctorate	0.320	0.289	0.223	0.254	0.225	0.226	0.281	0.293	0.236
	(0.382)	(0.382)	(0.377)	(0.390)	(0.381)	(0.379)	(0.379)	(0.382)	(0.380)
Contested electoral districs	-2.597***	-2.476***	-2.549***	-2.474***	-2.507***	-2.528***	-2.496***	-2.494***	-2.463***
	(0.487)	(0.487)	(0.483)	(0.491)	(0.486)	(0.482)	(0.483)	(0.485)	(0.484)
Second votes 2013	0.979***	0.973***	0.968***	0.975***	0.972***	0.976***	0.976***	0.975***	0.973***
	(0.0223)	(0.0226)	(0.0223)	(0.0225)	(0.0226)	(0.0221)	(0.0223)	(0.0224)	(0.0224)
Turnout	-0.107*	-0.106*	-0.119*	-0.106	-0.103	-0.0870	-0.0934	-0.105	-0.102
	(0.0640)	(0.0645)	(0.0640)	(0.0651)	(0.0648)	(0.0640)	(0.0645)	(0.0644)	(0.0641)
Business tax revenues (per capita)	0.368	0.299	0.376	0.287	0.336	0.498	0.379	0.281	0.265
	(0.810)	(0.816)	(0.811)	(0.818)	(0.820)	(0.812)	(0.814)	(0.813)	(0.811)
Business registrations/deregistrations	-0.0213	-0.0748	-0.0680	-0.0488	-0.0272	-0.0732	-0.0407	-0.0526	-0.0359
(per 1000 persons)	(0.181)	(0.182)	(0.179)	(0.181)	(0.181)	(0.180)	(0.180)	(0.180)	(0.180)
Unemployment rate	-0.0549	-0.0833	-0.0826	-0.0719	-0.0735	-0.0557	-0.0601	-0.0763	-0.0682
	(0.103)	(0.104)	(0.102)	(0.103)	(0.103)	(0.102)	(0.103)	(0.103)	(0.103)
Percentage male population	-0.541	-0.493	-0.603*	-0.530	-0.530	-0.502	-0.519	-0.535	-0.552*
	(0.331)	(0.335)	(0.331)	(0.335)	(0.334)	(0.330)	(0.331)	(0.333)	(0.332)
Senior-to-youth-rate	-1.875	-1.842	-2.046*	-1.928	-1.762	-1.919	-1.883	-1.915	-1.826

	(1.208)	(1.224)	(1.204)	(1.223)	(1.217)	(1.213)	(1.211)	(1.215)	(1.214)
Interactions incumbency_BT									
Incumbency_BT (runner up) x attractiveness	4.361** (2.165)								
Incumbency_BT (none of the cand.) x attractiveness	0.739 (1.783)								
Incumbency_BT (both cand.) x attractiveness	2.134 (1.612)								
Incumbency_BT (runner up) x competence		-0.847 (3.636)							
Incumbency_BT (none of the cand.) x competence		3.377 (3.436)							
Incumbency_BT (both cand.) x competence		0.955 (2.556)							
Incumbency_BT (runner up) x likability			5.406 (3.426)						
Incumbency_BT (none of the cand.) x likability			-0.754 (2.381)						
Incumbency_BT (both cand.) x likability			4.970** (2.385)						
Interactions gender									
Winner: female & Runner up: female x attractiveness				-0.338 (3.155)					
Winner: male & Runner up: female x attractiveness				-0.661 (1.663)					
Winner: female & Runner up: male attractiveness				0.00229 (2.416)					
Winner: female & Runner up: female x competence					0.567 (4.453)				
Winner: male & Runner up: female x competence					3.402 (2.506)				
Winner: female & Runner up: male competence					3.951 (3.046)				
Winner: female & Runner up: female likability						8.180** (3.862)			
Winner: male & Runner up: female likability						-1.599 (2.116)			
Winner: female & Runner up: male likability						-3.052 (2.879)			
Interactions age									
Age difference x attractiveness							-0.0650 (0.0487)		
Age difference x competence								-0.0311 (0.0739)	
Age difference x Likability									-0.0773 (0.0677)

Constant	38.32**	36.03*	42.93**	37.34*	37.92*	34.15*	35.65*	37.69*	38.05**
	(19.10)	(19.28)	(19.14)	(19.36)	(19.34)	(19.05)	(19.14)	(19.22)	(19.14)
Observations	259	259	259	259	259	259	259	259	259
R-squared	0.948	0.947	0.948	0.947	0.947	0.948	0.947	0.947	0.947

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1