Attractiveness, competence or likability?
Appearance effects in the 2013 German election

Sebastian Jäckle & Thomas Metz
Department of Political Science, University of Freiburg

Paper prepared for the ECPR General Conference in Montreal, August 26th – 29th 2015
Section: The study of Elections: Comparative Perspectives, Methodological Challenges
Panel: Personality and Voting

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 competence vis-à-vis his or her closest contestant significantly and substantially increases chances of winning a direct mandate. 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 analyzing 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.

Contact:
Sebastian Jäckle
Department of Political Science
University of Freiburg
Werthmannstr. 12
79098 Freiburg, Germany
Email: sebastian.jaeckle@politik.uni-freiburg.de
http://www.sebastianjaeckle.de
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 a more affectionate care by their mothers (Langlois et al. 1995), good-looking students receive better grades at school (Dukake et al. 2012) and have a higher likelihood of obtaining a college degree later on (Gordon et al. 2014). Attractive persons have better chances to get a callback when applying for a job (Bóo et al. 2013), they get paid better (Hamermesh and Biddle: 1994)¹ 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 has a detrimental effect 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 a number of controls (e.g. incumbency status) and confirmed a general effect of beauty in systems as diverse as Australia, Finland, Brazil or Mexico (King/Leigh 2009, Berggren et al. 2010, Lawson et al 2010) and also in Germany (Klein/Rosar 2005, 2010). 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/Rosenthal 1992) of information about the contestants to infer personal traits relevant to them which they cannot readily learn about otherwise (at least not at an acceptable cost). When trying to infer these attributes, candidate appearances such as attractiveness are usually easy to assess. This creates a halo effect in which the readily

¹ This is not only true for occupations in which beauty is obviously of major relevance (e.g. for escorts [Edlund et al. 2009]) but also for jobs in which “hotness” is generally not regarded as a necessary precondition for high salaries (e.g. university professors [Sen et al. 2010]).
available perception of beauty translates to other attributes and outshines other markers that might alternatively serve as a basis for judgment. 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. Studies even show that voters do not only resort to this kind of visual information short cuts in situations of information scarcity, but also when other types of information which would be more relevant for a rational and deliberative decision – e.g. information given in a newspaper article about the candidate – were readily available (Barrett/Barrington 2005).

Obviously, other possibilities to obtain information about candidates are available such as race, gender (McDermott 2009), or party (e.g. Dancey/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. 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 and a general increase of the role of the executive in parliamentary systems has led to the concern that politics could undergo a process of personalization. This could lead to a mismatch between a public focused on persons on the one hand and parliamentary systems usually awarding mandates to parties on the other (see Wagner/Weßels 2012: 72).

Clearly, 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 meager and that candidate effects do not change much over time (Wagner/Weßels 2012, Pappi/Shikano 2001, Gabriel et al. 2009, Ohr/Klein 2013), the picture is not yet complete since most work focusses on chancellor candidates, neglecting the level of constituencies. Here, work by Rosar and Klein (Rosar/Klein 2010, Rosar et al. 2008, Klein/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.

---

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.
such as perceived likability or inferred competence (but see Klein/Rosar 2005). Furthermore, although it is known that the perception of attractiveness is affected by available alternatives (Kenrick and Gutierres 1980), which implies that beauty is in part relative, studies on appearance effects in politics virtually all rely on absolute measures of attractiveness, modeling voter decision as an absolute assessment of candidate looks (much akin to a jury in a beauty contest) instead of more realistically capturing it as a relative decision between two or more available candidates who do not exist independently of each other.

In this article we will join and extend this line of research 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.

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\(^3\), 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 colors 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/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/Todorov 2010: 92]), but in the end it is perceived competence which is decisive. Expressed in terms of the social-psychological model of voting (Campbell et al. 1960), the difference between both perspectives is mainly 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

---

\(^3\) One of the very few studies finding no appearance effect comes from Bull et al. (1983). Yet, they argue that the landslide election victory of the Conservative Party in 1979 undermined the analysis as there were in the end no marginal constituencies where Bull and his colleagues supposed to find an effect of the candidate’s physical appearance.
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 – 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). Recently, Praino et al. (2014) broke new ground by arguing that whether competence or attractiveness is more important depends on the contestants’ gender: competence is more important for inter-gender races while in intra-gender races, it is attractiveness. They explain this in the following way:

“Between 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).

Others suggest that the effects of appearance are conditioned by several interactions:

1) 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/Leigh 2009: 591]).

2) 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/Leigh 2009: 592).

For somewhat different results see Poutvaara et al. (2009: 1134) who find Finnish female parliamentary candidates to profit from perceived beauty, but not from perceived competence, while exactly the opposite is true for their male counterparts.
3) 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/McKelvie 1993).

We test these three interactions, both for attractiveness and competence (and also for likability, see below). Other studies suggest further interactions which we cannot test but these should still be kept in mind when interpreting our results:

4) Appearance based effects should be stronger in low salience/low information elections than in elections with a high saliency because voters have less information to rely on (Lawson et al. 2010: 565).5

5) The effect of beauty is stronger in constituencies where many voters are not interested in politics in general, are not interested in the specific election and do not care who wins (King/Leigh 2009: 592).

6) The appearance advantage is more pronounced for voters with both low political knowledge and high TV consumption6 while watching a lot of TV does not intensify the effect of appearance for those more knowledgeable (Lenz/Lawson 2011: 586).

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, Doherty/Gimpel 1997), and once for Germany where Debus found an effect of both candidate and party likability (Debus 2010), 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

5 Studies from elections to an Australian local council (Martin 1978) and British community partnership boards (Banducci et al. 2008) for example detect appearance effects which are several times higher than those known from high salience elections, like those for the U.S. Congress.

6 Lenz and Lawson show that this effect is not only significant but also relevant: According to their estimation a one standard deviation increase in perceived appearance leads to almost 10 percentage points more votes for the respective candidate if the voter is from the bottom quartile of political knowledge and watches a lot of TV; if the voter has instead an average political knowledge watching a lot of TV only leads to an increase of 2.6 percentage points in the popular vote. According to its magnitude, the low knowledge/high TV effect would be comparable to the effect of newspaper endorsements or incumbency (Lenz/Lawson 586).
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. Analyzing 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/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.

3. Research Design

This section introduces our research design. First, we explain why the elections to the German Bundestag provide a good basis for the analysis of appearance effects. We then describe how we measured perceived attractiveness, competence and likability by means of an online survey. The third subsection explores which apparent features of the pictured faces foster the perception of the three traits. In subsection four we describe our weighting procedure for ambiguity in the ratings based on latency times and the fifth subsection gives an overview of the dependent variable and the other controls.

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

Many of the existing works studying the effects of perceived appearance focus on the United States where personalization is comparatively strong and parties relatively weak. Both conditions make it likely to find appearance effects. If, however, a “beauty premium” is universal, we should also be able to find it in a country where the role of parties is stronger and politics are less personalized. Following the idea of a least likely design (Eckstein 1975) such a finding would further strengthen the generalizability of the physical appearance effect on electoral success. Germany is well suited for such a tough test since it has strong and disciplined political parties which are the main focal points for voting decisions. Furthermore,

7 The two factors are derived from a principal components analysis. Dominance and facial maturity load high on “power” while likability and trustworthiness load high on “warmth”.

7
it is large enough (in terms of electoral districts/candidates) to allow for a meaningful statistical analysis. In German politics, personalization has somewhat increased during the last decades, yet 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/Shikano 2001, Brettschneider 2002, Klein/Rosar 2013). Furthermore, party identification is still strong in Germany and – as in other European parliamentary democracies – has a bigger influence on the final voting decision than in the presidential system of the USA (Berglund et al. 2005: 105-106). Most students of the German political system would therefore agree that it is still the parties which shape federal elections and not candidate personalities.

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 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.8

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 official websites of the candidates. Sometimes we also resorted to newspaper articles, or the Bundestags- or parliamentary party group’s official homepages. The idea behind that approach was to collect pictures that resemble those pictures used in the election campaign as much as possible since it is these pictures that voters will most probably have in mind when thinking of the candidate. 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 and scaled them to

---

8 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.
the same size. To prevent raters from associating the pictures with certain parties all party symbols were removed (e.g. lapel pins) and the pictures were transformed to a grey scale.\textsuperscript{9}

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. Our measurement thus differs from other studies which usually ask participants to rate all candidates and to rate them on an absolute scale. The idea of comparing only the two main alternatives draws on the observation by Kenrick and Gutierrez (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 facial looks of most candidates outside their district. Thus presenting raters with alternatives that may affect them 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.\textsuperscript{10}

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 non-included third candidate might still affect voters’ assessments. However, the problem is probably not be 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.\textsuperscript{11} And

\textsuperscript{9} Many of the SPD candidates use a red-violet corporate identity background which can no longer be identified when using a grey-scale photograph. The transformation seems unproblematic since it has already been shown that ratings of persons do not depend on the question whether the picture is in grey-scale or color (Klein/Rosar 2005: 279).

\textsuperscript{10} Praino et al. (2014) use absolute measures of attractiveness and competence. Yet, to account for the relational argument they calculate an attractiveness and competence differential.

\textsuperscript{11} In more than 90 percent 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.
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.

There is one general problem all studies on the effects of personal traits have to deal with: When raters recognize the candidates, their ratings will very likely not only be based on physical appearance, but also on personal fondness (or discontent) for the actual person or the candidate’s party. Different research strategies have been developed to avoid such a bias (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; Lewis & Bierly, 1990; Rosenberg 1991, 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) even go one step further. They present children and adults from Switzerland with photographs of candidates for the French National Assembly. Asking them whom of these persons they would prefer as a captain on a boat journey shows that the answers of the children were as good as the answers of the adults in predicting which of the candidates actually won in the election.

We take a different, two-stage approach. First, we excluded well known politicians from the sample with the help of a pre-test.12 Excluding major political actors which are known from nation-wide media is not only a precondition for obtaining unbiased ratings of perceived attractiveness, competence and likability, solely based on the presented photos, but can also be justified with respect to the underlying election model. For well-known candidates such as Angela Merkel or Minister of Finance Wolfgang Schäuble, voters can build on much more information than for the average candidate when making their decision. Appearance based effects should thus be smaller for them as the need to use information shortcuts decreases with the amount of information available about a candidate. On the basis of the pre-test we

12 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. We found that a mere positional approach – e.g. excluding all cabinet ministers – would not work properly in Germany: The pre-test showed that some cabinet members (such as the Federal Minister of Economic Cooperation and Development Gerd Müller) were virtually unknown whereas everybody seemed to be familiar with certain other politicians without any official or party-position (such as Hans-Christian Ströbele from the Greens). A full list of all excluded candidates can be found in the online Appendix O1.
excluded 40 of the 299 electoral districts. Furthermore, we asked the raters in the main survey whether they recognized any of the presented persons and asked for their names. Districts in which a candidate has been recognized by a rater in that way were also excluded from the analysis for that rater.\textsuperscript{13}

**Figure 1: Screenshot of online rating tool**

![Screenshot of online rating tool](image)

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. Candidate pairs presented to a rater were chosen randomly from the entirety of all 259 constituencies included in the survey. 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. For likability and competence pictures on the left were clicked slightly more often, yet this effect is very weak (see table A1).

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

\textsuperscript{13} Of the 6500 pairs of candidates presented, participants recognized one of the politicians in only 34 instances. These instances involved 29 different politicians in total, showing that no politician was recognized systematically, suggesting that the pre-test did a good job in filtering out all well-known candidates.
number to help participants remain concentrated during the whole process. 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 ten second countdown indicated to participants 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, 259 students from the University of Freiburg completed the survey which means that on average each electoral district was presented to 25.11 (sd = 4.93) raters. For attractiveness, the mean number of ratings for each pair of candidates is 22.04 (sd = 4.63; min = 11; max = 35), for competence it is 22.89 (sd = 4.68; min = 10; max = 35) and for likability it is 23.07 (sd = 4.67; min = 12; max = 35) 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 small number of raters can produce an accurate measurement of attractiveness (for example, King and Leigh (2009) have found that as few as four raters may produce a stable attractiveness results; Hamermesh and Biddle (1994) even found that a single rating by an interviewer yielded significant results). Before the start of the survey, participants were explicitly instructed to base their rating solely on their very first impression and their gut feeling. The instruction was based on the idea that according to Ballew and Todorov (2007: 17948), “rapid and unreflected face judgements” do a better job at predicting the electoral outcome (of gubernatorial and Senator races in the US) than when the participants are asked to make a deliberated decision (“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). The countdown reported above was chosen to support a quick decision; in practice, having to compare two pictures and having to click three buttons, this time span worked out very well. We measured not only which 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. These latency times are used 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, we test which properties of the candidates’ faces are adjuvant for being rated as the more attractive, competent or likable of the two presented candidates. We estimate six different models for each trait, one for each combination of sex of the rater and the three possible types of races (male-vs.-male/female-vs.-female, and female-vs.-male). Figure 2
presents the results. Most of them are fairly expectable: Age has a negative effect on attractiveness in all models, while it is only positively correlated with competence in male-vs.-male races. It has no significant effect on likability. Glasses tend to make people appear more competent while negatively impacting the other two traits. The opposite is true for beards. Wearing a suit or a blazer helps in female-vs.-female races as well as in male-vs.-female 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). A short hairstyle for women (virtually all men had short hair) is detrimental to being perceived as attractive while it has no effect on the perception of competence or likability. In male-vs.-male races, a bald head enhances perceived competence but depresses attractiveness or likability. The only measurable effect of jewelry is not the one potentially intended: 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 being more attractive or likable than women, but there is no difference in perceived competence. While all these findings are interesting as they show that there are certain apparent properties of facial pictures that shape our perception of the depicted person (although in some cases conditioned by the gender of the contestants), the most important finding for this study is that there are virtually no differences between the ratings of female and male raters. The only differences by trend are highlighted in figure 2 in black (e.g. when comparing two women, men rate older women as significantly less attractive, while this is not true for female raters) – yet all these little differences between male and female raters are not significant. 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

Female Rater

<table>
<thead>
<tr>
<th>Variable</th>
<th>Female vs. Female</th>
<th>Male vs. Male</th>
<th>Female vs. Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (in decades)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glasses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suit/Blazer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jewelry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short hairstyle for women</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tie</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bald head</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Male Rater

<table>
<thead>
<tr>
<th>Variable</th>
<th>Female vs. Female</th>
<th>Male vs. Male</th>
<th>Female vs. Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (in decades)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glasses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suit/Blazer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jewelry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short hairstyle for women</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tie</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bald head</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Odds ratio with 99% CIs

Likability ▼ Competence △ Attractiveness
3.4. Ambiguity in the ratings

Up to this point, we have interpreted all ratings for a candidate in the same way. Each click had the same weight. Yet, deciding who is more attractive, likable, or competent may sometimes be difficult to decide – 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 argument can be adapted to our context: Assuming that raters take more time to decide when in doubt, 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 two 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 between two candidates 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 = .33 to .51), suggesting that they tap into a common underlying ambiguity. Apparently, it makes sense to weight ratings with respect to latency times. Table O2 in the online appendix furthermore shows that randomization of buttons has no big influences on latency times. Of course, raters are different regarding their clicking-speed. For this reason we center 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). Correlating the final weighted scores for the

\[ \text{The final weighted proportion of clicks for candidate A in district X computes as follows:} \]

---

14 See Fazio et al. (2000) for a more thorough explanation why latency times are a good proxy for ambiguity in ratings. In a nutshell, the core idea is that a short latency time is due to a short and conflict-free activation process in the rater’s memory.

15 For the other two traits, likability and competence, the correlations are weaker but still there is some incidence that the three measures for ambiguity all measure at least somewhat the same.

16 The final weighted proportion of clicks for candidate A in district X computes as follows:
winning candidates of all electoral districts shows a medium strong relation ($r = .66$) between attractiveness and likability while the other two correlations between competence and attractiveness and competence and likability are below $r = .15$.

### 3.5. Dependent variable and controls

To analyze the effects of perceived attractiveness, competence, and likability on the performance of direct candidates at the 2013 general election, we estimate OLS regressions. As dependent variable, we use the differences between the vote shares for the two candidates of one electoral district which had been rated in the online survey, i.e. the winning candidate and the runner-up. On average, a winning candidate has a lead over the second-placed of 17 percentage points. The maximum difference is 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 and gender of the candidates, 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. Furthermore, we control for candidates holding a doctorate, as studies have shown a positive effect of an academic title in German elections (Schneider/Tepe 2011, Manow/Flemming 2011). Following our research questions, we also test interactions between the three trait variables and gender of the electoral race, incumbency and age in a second step.

### 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 the way we measure the incumbency status in the electoral district. In model 1 only those candidates

$$p_{AX} = \frac{\sum_{i=1}^{n_X} r_i(A)_X w_{i,X} }{\sum_{i=1}^{n_X} w_{i,X} }$$

with: $n_X =$ number of ratings in district $X$; $w_{i,X} =$ weight for rating $i$ in district $X$. $r_i(A)_X =$ rating $i$ for candidate $A$ in district $X$ (i.e. 0 or 1);

17 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

18 Gender of the electoral race is operationalized using four dummies: male/male; male/female; female/female; female/male (winner/runner up). Incumbency is coded in a similar way: winner is incumbent (has already won the most first votes in the 2009 general election); runner up is incumbent; neither winner nor runner up is incumbent. The age variable is measured as the difference in age between the winner and the runner up. A value of 10 for example means that the winner is ten years older than the second placed candidate.
who had won the first vote in 2009 are regarded as incumbents. There are thus three possible incumbency statuses: a) winner is incumbent (reference category in the model), 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 all candidates that had been members of the Bundestag during the last election period from 2009 to 2013 (hence four dummies can be created). Model 2 replicates model 1 using a stepwise procedure (the same applies to models 4 and 3 respectively).

Generally speaking, the explained variance is high and the results are consistent across the models.19 Regarding the three traits perceived attractiveness, competence and likability there is a clear finding: Being attractive and competent helps the candidate while likability does not have a significant effect. With the difference in the first votes between the winner and the runner up as dependent variable, the coefficients have to be interpreted in the following way: A change in the attractiveness value 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 2.128 percentage points (model 1). For competence, the effect is slightly stronger. 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 online appendix O3). Likability only becomes significant when attractiveness is omitted.

Gender plays a minor role: Only in electoral districts where the winning candidate is a man and the runner up is a woman, the difference between the two 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 electoral districts that had a priori been classified as contested. Interestingly, the difference in age of the candidates is only slightly significant in models 3 and 4. By trend, the older the winning candidate compared to the runner up, the smaller is her lead. The following interaction models investigate this result further.

19 Tests of multicollinearity showed no major problems (max. VIF = 5.1). According to Breusch-Pagan tests, only model 4 exhibits some heteroscedasticity. This model is thus 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.
### Table 1: Main effects models

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2 (stepwise)</th>
<th>Model 3</th>
<th>Model 4 (stepwise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived physical appearance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attractiveness</td>
<td>2.142** (0.884)</td>
<td>2.402*** (0.597)</td>
<td>1.572* (0.843)</td>
</tr>
<tr>
<td>Competence</td>
<td>2.389** (1.050)</td>
<td>2.529*** (0.916)</td>
<td>2.092** (1.000)</td>
</tr>
<tr>
<td>Likability</td>
<td>1.031 (1.065)</td>
<td>1.301</td>
<td></td>
</tr>
<tr>
<td>Incumbency dummies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incumbency_BT (only runner up)</td>
<td>-4.729*** (0.664)</td>
<td>-4.471*** (0.774)</td>
<td></td>
</tr>
<tr>
<td>Incumbency_BT (both cand.)</td>
<td>-2.583*** (0.468)</td>
<td>-2.642*** (0.484)</td>
<td></td>
</tr>
<tr>
<td>Incumbency_BT (none of the cand.)</td>
<td>-2.310*** (0.537)</td>
<td>-2.123*** (0.475)</td>
<td></td>
</tr>
<tr>
<td>Incumbency_direct (runner up)</td>
<td>-3.947*** (0.792)</td>
<td>-3.593*** (0.766)</td>
<td></td>
</tr>
<tr>
<td>Incumbency_direct (none of the cand.)</td>
<td>-1.956*** (0.500)</td>
<td>-1.515*** (0.464)</td>
<td></td>
</tr>
<tr>
<td>Gender dummies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winner: male &amp; Runner up: male</td>
<td>1.164 (0.811)</td>
<td>1.058 (0.771)</td>
<td></td>
</tr>
<tr>
<td>Winner: male &amp; Runner up: female</td>
<td>1.731* (0.898)</td>
<td>1.434* (0.853)</td>
<td></td>
</tr>
<tr>
<td>Winner: female &amp; Runner up: male</td>
<td>1.574* (0.912)</td>
<td>1.247 (0.861)</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age difference</td>
<td>-0.0209 (0.0176)</td>
<td>-0.0369** (0.0169)</td>
<td>-0.0391** (0.0159)</td>
</tr>
<tr>
<td>Doctorate</td>
<td>0.00341 (0.396)</td>
<td>0.194 (0.376)</td>
<td></td>
</tr>
<tr>
<td>Contested electoral district</td>
<td>-2.457*** (0.511)</td>
<td>-2.469*** (0.486)</td>
<td>-2.507*** (0.481)</td>
</tr>
<tr>
<td>Second votes 2013</td>
<td>0.975*** (0.0234)</td>
<td>0.982*** (0.0178)</td>
<td>0.976*** (0.0221)</td>
</tr>
<tr>
<td>Turnout</td>
<td>-0.105 (0.0672)</td>
<td>-0.0986 (0.0636)</td>
<td></td>
</tr>
<tr>
<td>Business tax revenues (in 1000 Euro per capita)</td>
<td>0.450 (0.847)</td>
<td>0.460 (0.804)</td>
<td></td>
</tr>
<tr>
<td>Business registrations/deregistrations (per 1000 persons)</td>
<td>-0.0291 (0.187)</td>
<td>-0.0673 (0.177)</td>
<td></td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>-0.0870 (0.108)</td>
<td>-0.0856 (0.102)</td>
<td></td>
</tr>
<tr>
<td>Percentage male population</td>
<td>-0.568 (0.345)</td>
<td>-0.543 (0.329)</td>
<td></td>
</tr>
<tr>
<td>Senior-to-youth-rate</td>
<td>-2.360* (1.256)</td>
<td>-2.167** (0.891)</td>
<td>-1.780 (1.198)</td>
</tr>
<tr>
<td>Constant</td>
<td>37.41* (19.92)</td>
<td>2.746** (1.203)</td>
<td>36.30* (18.92)</td>
</tr>
<tr>
<td>Observations</td>
<td>259</td>
<td>259</td>
<td>259</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.942</td>
<td>0.939</td>
<td>0.948</td>
</tr>
</tbody>
</table>

Standard errors in parentheses (Model 4 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 the online appendix (Table O5). For interactions with the metric variable “age difference”, significance is assumed if there is a marginal effect significantly different from 0 somewhere 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 and gender dummies 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

<table>
<thead>
<tr>
<th></th>
<th>Attractiveness</th>
<th>Competence</th>
<th>Likability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incumbency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winner</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Runner up</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No one</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female vs. female</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Male vs. male</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Female vs. male</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Male vs. female</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Age difference</strong></td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
</tbody>
</table>

“+” 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 won the first votes in 2009, attractiveness plays a major role. This means that attractiveness is more important for candidates that were new to the contest in 2013 and did not have the incumbency advantage. For the other two incumbency dummies “winner is incumbent” and “no one of the candidates is incumbent”20 there is no significant interaction effect. When the 2013 winner also won in 2009, the attractiveness effect cannot be distinguished from zero (the same is true for electoral districts where none of the candidates is an incumbent).

20 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”.
The second significant interaction effect is between gender and likability. Likability does not show any effect in electoral 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.\textsuperscript{21} 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 extent from their physical appearance has to be denied. Attractiveness and competence both show significant effects regardless of the candidates’ gender.

\textbf{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 in cases 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. Similar effects can be seen for competence, albeit to a lesser extent. For likability the interaction falls short of significance.

**Figure 5: Interactions between age difference and a) attractiveness, b) competence, and c) likability**

![Graph showing interactions between age difference and attractiveness, competence, and likability.]

5. **Discussion**

We started this article with the general question whether perceived attractiveness, competence or likability of direct candidates to the German Bundestag can serve as relevant predictors for their electoral success and which of these traits (if any) matters most.

To answer this question, we conducted an online survey among more than 250 students. 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. We also measured the latency
times, the raters needed to take their decisions. Using these latency times as an indicator of ambiguity, we aggregated the ratings at the level of the electoral district which gives us a measure of the perceived attractiveness, competence and likability of the winning candidate relative to the second placed. The relative measurement of the appearance based traits distinguishes this article from other approaches which measure perceived beauty (or other traits) in an absolute way (e.g. Rosar 2002). We argue that the relative measurement is better suited to replicate the actual situation within the electoral district. If physical appearance matters at all, voters will probably contrast the available candidates within a district with each other, but not with all other candidates in the rest of Germany, as the absolute measurement indirectly implies.

As an intermediate step of the 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. Interestingly, the effects are virtually the same for male and for female raters and they also differ only marginally with regard to the type of electoral race (male-vs.-male; female-vs.-female or female-vs.-male). Yet, these findings would be highly 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 in the main analysis that attractiveness and 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 theoretical discussion, 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 more important factor. Rather, both attributes seem to work independently of each other and their effects add up. From a more general vantage point, the relative measurement of appearances is not only more closely aligned with the factual situation in a constituency. Rather, our study has also shown that it is feasible, yielding sound measurement. In general, thus, our study 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 also plays a part although 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. Such interactions had been 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 studies (King/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, competence or likability compared to their opponents than older candidates.

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” (Hamermesh 2011) and is never “beastly” (Heilmann & Saruwatari 1979). 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. This might lead to a selection situation where it is perhaps even more important for a future candidate to be young and attractive, or to wear glasses making him appear competent than being qualified for the job or having a certain ideological position.
Acknowledgements

We thank all participants of the pre-test and the online survey and Stefan Wurster for establishing contact with the Heidelberg students.

Appendix

Table A1: Influence of left-right placement on the rating

<table>
<thead>
<tr>
<th></th>
<th>More attractive</th>
<th>More likable</th>
<th>More competent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pictures left</td>
<td>2855 (49.99%)</td>
<td>3049 (51.02%)</td>
<td>3017 (50.89%)</td>
</tr>
<tr>
<td>Pictures right</td>
<td>2854 (50.01%)</td>
<td>2927 (48.98%)</td>
<td>2911 (49.11%)</td>
</tr>
<tr>
<td>Phi</td>
<td>.0002</td>
<td>.0204</td>
<td>.0179</td>
</tr>
</tbody>
</table>


