Research Article

Social Perception of Male Faces. Individuals, Clusters, Dimensions

Ronald Henss[1]

1. Independent researcher

In an online experiment in a German and an English version, portrait photographs of 59 men from the age range late 20s to late 50s served as stimulus material. 1618 participants each rated a randomly selected photo on 36 personality-descriptive rating scales and estimated the age, body height and weight. In supplementary assessments, the degree of smile, degree of hair loss, hair color, facial hair, glasses, formal dress, head rotation, head tilt and brightness of the image background were determined. All variables show sufficient, mostly high to very high reliability. The data analysis was carried out at the level of the stimulus persons, on the one hand from a dimensional and on the other hand from a typological perspective. A principal component analysis of the personality-descriptive traits yielded a five-dimensional space with orthogonal factors. These can be interpreted straightforwardly as Social Agreeableness, Attractiveness, Masculinity, Status/Intelligence and Negative Affects. Using multiple regression, the age estimates and the degree of smiling can be fitted into the psychological space in an excellent way. Wearing glasses and height estimates can also be localized well in this space. In addition to the customary dimensional approach, a typological approach was adopted, which is rarely found in this field of research. In a hierarchical cluster analysis of the five personality dimensions and the external characteristics, groups of stimulus persons were identified who are similar to each other in a bundle of characteristics and who differ markedly from other groups. These clusters were mapped into the five-dimensional face perception space. In this way, it is demonstrated that not only does each approach provide interesting insights in its own right, but that both approaches mutually enrich each other by combining them. A main emphasis of this paper is the comparison with the two-dimensional Valence-Dominance model, which has been the most influential paradigm for more than a decade. It is not disputed that evaluation on the Badness – Goodness dimension is of paramount importance, nor is the importance of Dominance or Power called into question. Nevertheless, it is shown that a two-dimensional model cannot do justice to the complexity of face perception. Face perception is much more differentiated and – this is the crucial point – it largely coincides with the personality structure known from differential psychology, which is based on self-assessments and assessments of acquaintances.

 doi.org/10.32388/6IDC9O

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

Introduction

The subject of this study is the assessment of strangers on the basis of facial photographs. Our starting point is our paper „Hair Loss, Body Height and Attractiveness Malus for Men“ (Henss, 2024a). The data set of that study is the basis of the current research. We will expand the data set and analyze it from different perspectives. The components of the title of the previous paper – hair loss, height and attractiveness – will still be considered, but they are only three variables among others.

Before we address our current research questions, we take a look at the starting point. Our central interest was the impact of genetically induced male pattern hair loss (androgenetic alopecia) on the beholder. For a review on this topic, see Henss (2001); for empirical work, see for example Becker (2003), Chan et al. (2019), Henss (2024a, b), Kranz, Nadarevic and Erdfelder (2019).

The study was conducted as an online experiment in a German and an English version. The participants were asked to evaluate a randomly selected portrait photograph with regard to various personality traits and to estimate age, height and weight. The photos were taken from a model catalog and accordingly the men, viewed as a group, were considerably more attractive than their peers. In an additional study, the degree of hair loss was assessed.

The first interesting finding relates to the absolute level of the assessments. With one exception, the men were rated quite favorably. The exception is attractiveness. Although there is no doubt that our specific sample of male models are more attractive than their peers, the attractiveness ratings are downright demeaning. The fact that men generally score poorly in attractiveness ratings has been repeatedly observed in our studies on facial judgments (Henss, 1992a, 1998a; for a counterexample, see Henss, 1987). This is also well known from other studies in attractiveness research. We used to call this phenomenon attractiveness bonus for young women. However, since young women are not considered particularly attractive on average, whereas the majority of men receive poor ratings, it is more appropriate to speak of an attractiveness malus for men.

A second interesting finding concerns the level of analysis. The data were analyzed on the one hand at the level of the judges and on the other hand at the level of the stimulus persons. At the level of the judges, both the degree of hair loss and the estimated age and height are important variables in the sense that they correlate with almost all other variables. At the level of the stimulus persons, however, hair loss only correlates with attractiveness and to a lesser extent with mood. Age and height, on the other hand, are of great importance at both levels of analysis.

The fact that age plays a prominent role in personality impressions is no surprise. If one considers the entire lifespan, there is no variable that brings about changes anywhere near as dramatic as age. In this sense, age is *the* ultimate super variable. But even the changes in the age range we are looking at, from the late 20s to the late 50s, entail profound changes to the face that affect personality impressions. On the paramount importance of age for facial assessment, see in particular the impressive research program by Leslie Zebrowitz (Montepare and Zebrowitz, 1998; Zebrowitz, 1997, 2011).

The fact that body height plays a very important role for men is generally known and empirically well documented. The difference in the average height of men and women is around two standard deviations (Henss, 2017; NCD, 2016). With regard to psychological characteristics, such a huge effect size is extremely rare. Height is an indicator of health, undisturbed growth, strength and intelligence, and being taller is an advantage for men in a variety of areas of life (on different aspects see for example Blaker et al., 2013; Case and Paxson, 2008; Harper, 200; Harris, Brett, Deary and Starr, 2016; Judge and Cable, 2004; Pisanski et al., 2022; Roberts and Herman, 1986; Tyrell et al., 2016; Vuoksimaa et al., 2018; Wang et al., 2020). Nevertheless, it is astounding that height plays such a substantial role in the assessment of headshots. Such photos hardly provide any direct cues for estimating height. And yet almost all participants are able to make such estimates and these estimates paint a consistent picture. With this observation, we come to the core of the present study.

Consistency and consensus in facial judgments

The human face is the most fascinating object in our lifeworld. Nothing evokes anywhere near as manifold impressions as the sight of a face. These impressions have the character of an instinct; they arise involuntarily, unconsciously and at lightning speed.

One of the most impressive findings of psychological research concerns the almost unbelievable ability to form stable impressions of a stranger even with minimal viewing time. In a number of studies, faces of strangers were presented for such a short time that they were barely perceptible, and yet subjects had no problems making judgments about various personality traits. Presentation times of 100 milliseconds or even less are sufficient for this. The crucial point is that these judgments are essentially the same as with any length of viewing time. For empirical studies, see for example Goldstein and Papageorge (1980), Locher, Unger, Sociedade and Wahl (1993), Bar, Neta and Linz, (2006), Willis and Todorov (2006), Borkenau, Brecke, Möttig and Paelecke (2009). In our laboratory, we have also conducted such experiments and shown that for different characteristics a presentation time of 300 or even only 150 milliseconds is sufficient to give intra-individually consistent and inter-individually concordant ratings (Henss, 1992a, p. 153; Kurz and Register, 1998; Schmidt, 1999).

Our judgments of unfamiliar faces are intra-individually stable, there is a certain consensus between individual judges, individual judges usually agree well to very well with the average group opinion and the agreement between group standards is often very high. As might be expected, the degree of agreement at the various levels depends on the characteristic being assessed. The different types and the degree of agreement in judgments of facial photographs was one of the focal points of our research (Henss, 1992a, 1997a and especially 1998a). Here are a few summary findings. Extraordinarily high concordance is found for age, especially when it is estimated numerically, but also for rating scales such as 'young – old'. Physiognomic characteristics are usually assessed with greater agreement than psychological characteristics, although interestingly, physiognomic characteristics that can be measured using a ruler score less well than, for example, ratings relating to hair, skin texture and eyebrows. What is interesting in relation to our original topic is that the scale 'Hair: sparse - full' shows particularly concordant ratings for men. Among psychological characteristics, the highest agreement is often found for positive mood, attractiveness and extraversion. Extremely low agreement is usually found for emotional stability and negative affects. These traits apparently requires dynamic information, such as that conveyed by moving facial expressions or voice (Borkenau and Liebler, 1992). In this study, we will report the degree of consensus among judges (also see, for example, Hehman, Sutherland, Flake and Slepian, 2017; Hester, Xi, Hehman, 2021).

Before we turn to our core topic, two comments are in order. First, while first impressions are consistent and stable, they are not carved in stone and can be modified by additional information. Second, this paper is not about the question of right or wrong. Our impressions can have a grain of truth, sometimes even a big one, but usually it is small and not infrequently zero. However, this does not change the fact that our impressions, regardless of whether they are right or wrong, have tangible consequences in various areas of real life.

Structure of personality impressions

Our focus is on another aspect. Our impressions when looking at other people's faces relate to a wide variety of features, but these impressions are not unconnected. Quite to the contrary, there are more or less strong correlations between different features and the pattern of these correlations leads to an orderly, coherent structure. The remarkable point is that this structure is often an impressively clear reflection of the personality structure that emerges in self-assessments or in the assessment of acquaintances.

As a starting point, we look at a section of Alexander Todorov's research program, which has arguably been the most influential paradigm for more than a decade. An excellent account can be found in Todorov and Oh (2021), for a broader overview for a wider audience see Todorov (2017). This research program is much more extensive and sophisticated than we can discuss here. We limit ourselves to a small segment that is pertinent to our topic.

The seminal paper is Oosterhof and Todorov (2008). The stimulus material was a set of 66 facial photographs of men and women aged 20 to 30 with neutral facial expressions. In the first step, these photos were presented to participants who were asked to write down their spontaneous impressions. The 1134 descriptions were grouped into trait dimensions, which were mentioned most frequently, and the category Dominance was added out of special interest. In the end, the following 13 items were considered: aggressive, attractive, caring, confident, dominant, emotionally stable, intelligent, mean, responsible, sociable, trustworthy, unhappy, weird.

In the second step, the photos were rated by another group according to these characteristics. A principal component analysis yielded a two-dimensional solution. The first component explains 63.3 percent of the variance, the second 18.3 percent. On the first component, which explains the lion's share of the variance, all characteristics that are generally considered positive have a positive loading and all characteristics that are considered negative have a negative loading. So this is definitely an Evaluation or Valence factor. As trustworthy has the highest loading, the authors call this dimension Trustworthiness. On the second component, the highest loadings are for dominant, aggressive and confident. Based on the highest loading, the authors refer to this dimension as Dominance.

These two basic dimensions are immediately reminiscent of Osgood's EPA model – E: *Evaluation*, P: *Power*, A: *Activity* – except that the Activity dimension is missing here (Osgood, 1969; Osgood, Suci and Tannenbaum, 1957). The model also shows a large overlap with Wiggins' *Interpersonal Circumplex* (Wiggins, 1979; Wiggins, Phillips and Trapnell, 1989; Wiggins and Pincus, 1992), which is spanned by the dimensions *Agency* and *Communion*, also known as the *Big Two* of the interpersonal domain. The 2D model is thus well integrated into the research landscape and, under the name of *Valence-Dominance model*, it has come to be the dominant paradigm in face assessment research.

As long as one limits oneself to this set of 13 items and analyzes the data using principal component analysis, the 2D model can claim almost universal validity.[1] Nevertheless, the model is hopelessly undercomplex for our research questions.

It is evident a priori that 13 items cannot be sufficient to capture the manifoldness of human personality and the multifaceted impressions when looking at faces. Todorov and Oh are well aware of this limitation and refer to their model as a simplistic 2D model. They are also aware that concentrating on the faces of young adults may conceal relevant aspects. In the following, we will only mention a few studies that offer a different perspective and then present some of our own research findings.

Sutherland et al. (2013) used 1000 photos of Caucasian adult faces from a wide age range, which, unlike in many other studies, showed the faces as they would be encountered in everyday life. In a test of Oosterhof and Todorov's 2D model, they obtained an additional factor of Youthfulness/Attractiveness. In a further

experiment with morphed images, they were able to replicate this finding. See also Vernon, Sutherland, Young and Hartley (2014).

In the study by Wolffhechel et al. (2014), a portrait photo was taken of each of the 244 participants and they were asked to rate each of twenty randomly selected unknown faces from this set with regard to adventurous, attractive, dominant, emotionally stable, extravert, friendly, intelligent, masculine, physically healthy, responsible and temperamental. The participants were also asked to assess their own personality on a self-assessment questionnaire. Based on the pattern of correlations between the facial judgments, the authors constructed three clusters, which they termed Trustworthiness-Friendliness, Attractiveness-Health-Extraversion and Dominance-Masculinity. While we would interpret the pattern slightly differently,[2] the key point here is that Attractiveness and Health and Sex Typicality are added, and that these are closely related. It should be noted in passing that for some traits, the facial assessments correlated with the self-assessments of the persons being judged (r between.20 and.32; p <.01).

The studies considered so far have the serious shortcoming that they included far too small a number of characteristics and thus had no chance of identifying a more differentiated personality structure. We are now looking at studies that have rectified this shortcoming.

Lin, Keleş and Adolphs (2019) had 50 male and 50 female photos rated on 100 English trait words and found four dimensions, which they labeled Warmth, Competence, Femininity, and Youth. This four-dimensional space was similarly replicated in North America, Peru, Latvia, the Philippines, Kenya, Gaza, and India. The authors explicitly point out that the use of 100 instead of only 13 items „not only revealed a larger number of dimensions but a dimensional space that is distinct from prior frameworks“ (p. 6). Although Warmth and Competence are semantically similar to the 2D dimensions Valence and Dominance, their content is shaped by differing attributes.

Walker and Vetter (2015) investigated the relationship between the Oosterhof-Todorov two-dimensional Valence-Dominance model and the Big Two of the social domain – Agency and Communion – and the Big Five factors Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness. The stimulus material consisted of 153 facial photographs of young adults with neutral expression. In an online experiment, 1671 participants each assessed three randomly selected faces. The Valence-Dominance model was represented by the items trustworthy and dominant, the Big Two and the Big Five by well-established questionnaires. The item trustworthy has an extraordinarily high correlation with Communion (.94) and is completely independent of Agency (.05). Dominant correlates with Agency at.66 and with Communion at -.64. A multiple regression with trustworthy and dominant as predictors explains 90 percent of the variance of Communion and 69 percent of Agency. The interpretation „the two basic dimensions of face evaluation and

the Big Two personality dimensions are not only semantically, but also empirically akin to each other“ (p. 8) holds without reservation for trustworthy, somewhat weaker for dominant. For the Big Five, trustworthy shows a high correlation with Openness (.87) and Agreeableness (.84) and dominant correlates with Agreeableness at -.73. The multiple regression on trustworthy and dominant yields the following percentages of variance explained: Agreeableness 81, Openness 76, Conscientiousness 57, Extraversion 47 and Neuroticism 46. Thus „it is evident that the Big Five capture more than the two basic dimensions of face evaluation“ (p. 13).

From the late 1980s to the early 2000s, we have investigated the question of dimensions of personality impressions in detail with different samples of adult faces of both sexes and from a wide age range, with different samples of participants, different characteristics, different scales, in different experimental setups and often in a German and an English online version. In the following, we provide a brief summary.[3] It should be noted that our analyses were conducted at the level of the judges and not at the level of the stimulus persons. At the level of the judges, there may be one or two more factors and the overall variance explanation is much smaller than at the level of the stimulus persons.[4] Depending on the research questions, we considered about 30 to more than 100 personality-descriptive items, mostly adjectives or short phrases, occasionally also type nouns (on type nouns see Henss, 1995, 1996, 1998b). Usually, our items were intended to capture the Big Five, various facets of physical attractiveness and aspects of mood. In principal component analyses with varimax rotation, we always obtained at least four clearly interpretable components and it was not uncommon for the number of eigenvalues > 1 to be greater than ten. Congruence analyses according to Tucker usually yielded six or seven reproducible components, and there were some remarkably stable findings across various studies under very different conditions. We almost always found three strong components, namely *Attractiveness*, *Extraversion/Mood*, and *Social Agreeableness*, whereby Extraversion was almost always coupled with *Positive Affects*.[5] These are our *Big Three of Face Perception*. In addition, there was almost always a component that we call *Self-Assurance*, which is marked by adjectives such as self-assured, confident, independent, strong, self-reliant, superior, insecure(-). Further differentiation resulted in a clearly circumscribed *Conscientiousness* factor and then an *Intellect/Openness* factor. An independent factor *Emotional Stability*, on the other hand, could only rarely be identified. For higher-dimensional solutions, we were able to identify *Sex Typicality*, *Fashion* and *Health* as smaller specific factors, which are otherwise mainly found on the Attractiveness factor.

With regard to the above considerations, it should be pointed out that we have never found a component that could be interpreted as Valence. Nor did we find a factor that had dominance as its core. We will elaborate on this in the discussion.

In addition, we sometimes used a list of 30 bipolar rating scales to capture physiognomic characteristics. This physiognomic differential never gave a meaningful factor structure. Today's methodologically and technologically much more sophisticated approaches have made it clear why such a „small" set of physiognomic rating scales is doomed to failure. The *data-driven computational model of social judgments* in Todorov's research program takes into account 100 or more components, each containing an impenetrably complex mixture of length and angle measures and brightness, saturation and hue. With the help of such a super-high-dimensional space, the *face perception space*, which is smaller by a multiple, can be understood in an unprecedented way (Todorov and Oh, 2021).

Aims of the present study

For the previous study (Henss, 2024a), whose data set we are now using, we selected items that were of interest with regard to our other studies on social perception of male pattern baldness. The aim here was not to cover a broad spectrum of the personality sphere and, in particular, it was not about a systematic consideration of the Big Five. We had analyzed the data at the level of the judges and the level of the stimulus persons. We are now concentrating on the stimulus person level. In doing so, we look at the data from two different perspectives.

First, we want to construct a simple low-dimensional space of psychological perception of faces in which the stimulus persons can be located. This corresponds to the conventional factor analytic approach described above.

In addition to the dimensional analysis, we apply a typological approach that is rarely found in research on face perception. The aim here is to identify and describe groups of persons whose members are similar to each other in terms of a bundle of characteristics and who can be clearly distinguished from other groups. For this purpose, we use cluster-analytical methods.

A central aim is to show that the two approaches can be combined in a fruitful way.

Subjective (psychological) and objectifiable (external) characteristics

The data set of the previous study comprises 36 personality-descriptive items, estimates of age, body height and weight and the body mass index BMI derived from these, as well as an assessment of the degree of hair loss. Here it is crucial to note an important distinction. The personality-descriptive items capture attributes that can only be assessed subjectively. Although age, height, weight and hair loss were also assessed subjectively, they could in principle be measured objectively. All these characteristics belong to personality in the broader sense. We will refer to the characteristics that can only be measured subjectively as psychological

or personality traits (in a narrower sense) and those that could be measured objectively as external characteristics.

Specifically for the current study, we collected additional data on external characteristics. These are hair color, facial hair, glasses, formal dress, smile, head rotation, head tilt and brightness of the image background.

We will use the subjective features to construct a psychological space of face perception. We will then explore how the external variables relate to this space. Next, using the dimensions of the psychological space and the external variables, we will identify groups of individuals that represent clearly distinguishable types. Finally, we will map these types in the psychological space.

Methods

Stimulus material

This study is based on black and white portrait photographs of 59 men taken from a model catalog. They represent different types, but, as the source reveals, they are not a representative sample of German men. On the contrary, taken as a group, they are undoubtedly considerably more attractive than the average of their peers.

Procedure

The study was conducted as an online experiment via our home page at the Psychological Institute of the Saarland University in Saarbrücken, in a German and an English version. The participants were asked to rate a single randomly selected photograph on a five-point rating scale with regard to 36 personality traits and to estimate the age, body height and weight.

In a supplementary study, 21 undergraduate students of psychology (15 women, 6 men) independently rated all 59 photos according to the degree of hair loss. As a yardstick, Norwood's (1975) well-established classification system was used, which distinguishes seven levels from a full head of hair (Type I) to full-blown androgenetic alopecia (Type VII).

The two studies were the basis of the previous paper and its results are the starting point for the current analysis.

For the current analysis, additional variables were taken into account. On the one hand, the 59 photos were independently assessed by a convenience sample of three men and two women according to the following characteristics: Hair color (white/grey, light, medium, dark, black), degree of smile (none, light, medium, wide), predominant side of the face („From which side of the face – as seen from your perspective – do you see

more?"; with the levels significantly more left, slightly more left, both equal, slightly more right, significantly more right) and brightness of the image background (white, light, medium, dark, black). On the other hand, we ourselves determined the characteristics of facial hair (none, moustache, beard, beard and moustache, moustache and lush full beard), glasses (no, yes) and tie (no, tie, bow tie) and measured the head tilt, i.e. the rotation towards the shoulders, using a protractor. Here we differentiate between the measured values, which also take the direction into account, and the absolute values.

Thus, in addition to subjective personality traits, we take into account a number of external characteristics of the stimulus persons and some characteristics of the shooting technique.

Variables and level and objective of analysis

The main study resulted in a data set of 1618 participants (1137 women and 481 men; English version 980, German version 638; age 14 to 67, mean 25.4). In the previous study, these data were analyzed in the first step at the level of the judges. In the second step, the data were analyzed at the level of the stimulus persons. This means that the unit of observation is the mean value of the respective characteristic.

We now continue the analysis at the level of the stimulus persons. From the previous study, we adopt the 36 personality-descriptive items, which were assessed on a 5-point rating scale, as well as the estimates of Age, Height, Weight and the derived Body Mass Index BMI (kg/m^2) and the degree of Hair Loss. In addition, we take into account the variables Hair Color, Beard, Glasses, Tie, Smile, Face Side, head Tilt, Tilt absolute and Background brightness.

First, we want to use factor-analytical methods to construct a low-dimensional space of personality impressions from the 36 personality-descriptive items and interpret this space psychologically. Then we want to explore the relationship between the external variables and this personality space. Secondly, we want to adopt a typological approach and use the personality factors and the external variables to identify groups of men who are similar to each other in a number of respects and who differ markedly from other groups. For this, we use cluster-analytical methods. Finally, we want to link the dimensional and typological perspectives and map the individuals and the clusters in the personality space.

Results

Consensus among judges, reliability of the group standards

Our first question concerns the reliability of our data. Two test arrangements must be distinguished. The main experiment was conducted as a single stimulus assessment. Here, each subject rated only a single

randomly selected face on the 36 personality-descriptive rating scales and estimated age and height and weight, from which BMI was derived. The remaining variables were obtained in a serial stimulus assessment, i.e. the subjects assessed all 59 men. There were 21 judges for the assessment of Hair Loss and 5 for Hair Color, Smile, Face Side and Background. The variables Glasses, Beard, Tie, Tilt and Tilt absolute were determined by ourselves. Here the characteristics are so evident that no reliability analysis is required.

In the case of single stimulus judgments as in our main experiment, the measure of agreement is the Intra-Class-Correlation ICC (Koo and Li, 2016; Shrout and Fleiss, 1979; Shrout and Lane, 2012). Two indices must be distinguished here, namely ICC(1,1) and ICC(1, k). The first parameter, in our case 1, denotes the statistical model, which in our case is a one-way analysis of variance without repeated measures (ANOVA). The second parameter refers to the unit of observation. ICC(1,1) refers to the individual level and corresponds to the proportion of variance attributable to the differences between the faces in relation to the total variance. This is the reliability at the level of the individual judges. With ICC(1, k), k denotes the number of judges per photograph and the value can be determined from ICC(1,1) by means of the Spearman-Brown formula. This index denotes the reliability of the group standard. Since we will analyze the data at the level of the stimulus persons, this is precisely the figure that is important for us. The intra-class correlations ICC(1,1) and ICC(1, k) from our main experiment are summarized in Table 1.

		ICC(1,1)	ICC(1, k)
in a good mood	gutgelaunt	.34	.93
merry	fröhlich	.31	.92
cheerful	heiter	.30	.92
sexy	sexy	.29	.92
seductive	verführerisch	.28	.92
good looking	gutaussehend	.25	.90
high occupational status	angesehener Beruf	.24	.90
successful with women	Erfolg bei Frauen	.23	.89
career oriented	karriereorientiert	.22	.89
erotic	erotisch	.21	.88
educated	gebildet	.21	.88
dominant	dominant	.18	.86
mature face	reifes Männergesicht	.17	.85
successful in his job	Erfolg im Beruf	.17	.85
good-natured	gutmütig	.17	.85
masculine appearance	typisch männlich	.17	.84
dangerous	gefährlich	.16	.84
aggressive	aggressiv	.15	.83
family oriented	familienorientiert	.15	.82
likeable	sympathisch	.14	.81
intelligent	intelligent	.13	.80
belligerent	angriffslustig	.12	.79
baby face	Babyface	.12	.79
likes children	kinderlieb	.12	.79
sincere	aufrichtig	.11	.77
sad	traurig	.11	.77

		ICC(1,1)	ICC(1, k)
honest	ehrlich	.11	.77
pronouncend male face	markantes Männergesicht	.10	.76
unpredictable	unberechenbar	.10	.75
naive	naiv	.06	.65
timid	schüchtern	.06	.64
nervous	nervös	.06	.63
insecure	unsicher	.05	.59
withdrawn	zurückhaltend	.04	.55
anxious	ängstlich	.04	.54
earnest	ernst	.03	.46
Age	Alter	.69	.98
BMI	BMI	.18	.86
Weight	Gewicht	.15	.83
Height	Größe	.12	.78

Table 1. *Consensus among judges. Single stimulus assessment.*

For the subjective attributes, the ICC(1,1) values range from.03 to.34, with a median of.15. The highest agreement is found for items relating to a positive mood, attractiveness and the professional sphere. There is strikingly little consensus on negative emotions and social withdrawal. Readers who are less familiar with this field of research may have the impression that the inter-individual consensus on subjective personality traits is astonishingly low, but this would be a mistake. We will come back to this in the discussion. Here we are interested in the reliability of the group standard ICC(1, k). On average, each picture was rated by 27.4 judges and the reliability ranges from.46 to.93 and the median is.83. Measured against the conventional minimum requirement of.70, the reliability is high to very high in the majority of cases, with only a few items

showing insufficient reliability. However, we will not look at the individual items, but rather aggregate them to personality factors, which by necessity have a higher reliability.

Among the external variables, Age stands out with an ICC(1,1) of.69 and almost perfect reliability of the group standard. The values for BMI, Weight and Height, on the other hand, correspond to the middle range of the subjective personality traits.

In the case of serial stimulus assessments, consensus is determined via reliability analyses. Here we distinguish between three levels. Firstly, the average correlation between two individual raters, which we refer to as r(i, j). This is the counterpart to ICC(1,1). Secondly, the correlation between the individual judgment and the average judgment of the rest of the group; r(i, rest). In test theory, this is the corrected item-total correlation. Thirdly, the reliability of the group standard, which is measured by Cronbach's α and McDonald's ω. The indices are summarized in Table 2. For r(i, rest) the maximum and minimum are indicated. N denotes the number of judges.

	r(i, j)	r(i, rest)	Cronbach's α	McDonald's ω	N
Hair Loss	.89	.87 –.97	.99	.99	21
Hair Color	.88	.85 –.90	.97	.96	5
Smile	.82	.85 –.94	.95	.96	5
Face Side	.69	.75 –.85	.92	.92	5
Background	.79	.90 –.93	.95	.95	5

Table 2. *Consensus among judges. Serial stimulus assessments.*

Now we see a completely different picture. Here, consensus is already very high at the level of the individual judges. The agreement of the individuals with the rest of the group is by necessity higher and the reliability of the group standard is extraordinarily high. For Hair Color, Smile, Face Side and Background, it should be pointed out that the extraordinarily high reliability was achieved with only five judges. The enormous differences between Table 1 and Table 2 are due on the one hand to the fact that the external features are much clearer to recognize than the subjective traits, and on the other hand to the fact that in the case of serial stimulus judgments the 59 photos provide a common frame of reference, whereas in the case of single stimulus judgments no common frame of reference is available.

doi.org/10.32388/6IDC9O

Since our following analyses are based on the group standards, we are throughout dealing with reliable, mostly very reliable and some extremely reliable variables.

The personality impressions space

In the main experiment, the stimulus persons were assessed on 36 personality-descriptive rating scales. Our first aim is to construct a space on the basis of these data and to interpret it psychologically and to localize the stimulus persons in this space.

For this purpose, we conducted a principal component analysis (PCA) with varimax rotation on the one hand and an exploratory factor analysis (EFA) with maximum likelihood method and oblimin rotation on the other hand. The trend of the eigenvalues, which can be seen in the first data column of Table 3, suggests a five-dimensional solution according to both the Kaiser criterion (eigenvalue > 1) and a scree test of the entire trend pattern. The second and third data columns show the percentage of variance explained for the components and factors respectively. The right-hand column shows the degree of agreement between PCA and EFA, measured by the correlation between the factor scores of the stimulus persons.[6]

		Percent variance		
Fact/Comp	**Eigenvalue**	**PCA**	**EFA**	**r(PCA;EFA)**
1	13.43	29.8	22.3	0.97
2	7.34	16.2	16.8	0.97
3	4.82	15.8	15.3	0.97
4	3.24	14.0	13.7	0.98
5	1.84	9.5	14.6	0.79
Total		85.2	82.7	

Table 3. *Principal component analysis and exploratory factor analysis.*

In the principal component analysis, the five components explain 85.2 percent of the total variance. We thus have an extraordinarily high exhaustion of information. As is usually the case with principal component analyses, the first component is particularly strong at 29.8 percent. The second, third and fourth components

are roughly equally strong and the fifth is somewhat weaker at 9.5 percent. The exploratory factor analysis explains 82.7 percent of the variance and the differences in the explained variance of the factors are smaller due to the method.

For the first four components/factors, the match is almost perfect (right column), but for the fifth, the correspondence is notably weaker. We will come back to this in a moment.

In the following, we will only consider the PCA. This explains a slightly higher proportion of the variance, but the main reason is orthogonality. The PCA spans a five-dimensional space whose dimensions are independent of each other. With EFA, on the other hand, the correlations between the dimensions must always be taken into account. In addition, EFA has stricter requirements than PCA.

Table 4 shows the factor loadings of the PCA and, in the right-hand column, the residual portion of the variance that is not explained by the five components (1 minus communality).

	1	2	3	4	5	res
good-natured	.93					.08
sincere	.89					.13
honest	.89					.17
likeable	.88					.13
likes children	.86					.19
belligerent	-.86					.20
dangerous	-.84			-.33		.14
aggressive	-.83		.33			.14
cheerful	.80				-.51	.05
in a good mood	.75				-.59	.07
unpredictable	-.74					.33
merry	.72				-.55	.10
dominant	-.71	.32	.52			.11
family oriented	.71	-.32		.36		.24
sexy		.96				.05
good looking		.94				.05
erotic		.93				.08
seductive	-.31	.92				.04
successful with women		.88	.39			.05
mature face			.92			.10
pronouncend male face		.40	.77			.18
anxious			-.77			.21
baby face	.35		-.76			.23
masculine appearance	-.36	.49	.72			.10
nervous		-.38	-.72		.33	.22
naive	.42		-.66			.31

doi.org/10.32388/6IDC9O

	1	2	3	4	5	res
insecure		-.46	-.57	-.31	.41	.20
successful in his job				.95		.04
high occupational status				.94		.07
career oriented				.93		.13
educated				.92		.08
intelligent				.85		.22
withdrawn	-.34				.72	.30
sad	-.49				.72	.18
earnest				.43	.70	.29
timid	.34		-.58		.60	.11

Table 4. *Principal component analysis. Factor loadings and communalities.*

The interpretation of the loading pattern is straightforward, especially for the second, third and fourth components.

The first component has, unsurprisingly, the largest number of high loading items as well as some non-trivial secondary loadings. The high loading items honest, likeable, likes children, belligerent(-), dangerous(-), aggressive(-), cheerful, in a good mood, unpredictable(-), merry, dominant(-), family oriented paint a picture of a socially agreeable man who is in a good mood and a good family father. We term this component *Agreeableness*.

The second component is defined by sexy, good looking, erotic, seductive, successful with women with very high loadings. We refer to it as *Attractiveness*, but it should be noted that the focus is on the sexual aspect. The secondary loadings of masculine appearance and pronounced male face should also be mentioned here, which point to a connection between male attractiveness and a masculine appearance.

Masculine appearance is the key aspect of the third component, whose major loadings are mature face, pronounced male face, anxious(-), baby face(-), masculine appearance, nervous(-), naive(-), insecure(-).[7] The secondary loadings of timid(-) and dominant should also be mentioned. It is reasonable to call this component *Masculinity*.

doi.org/10.32388/6IDC9O

The fourth component, constituted by successful in his job, high occupational status, career oriented, educated and intelligent, is a mixture of an orientation towards the occupational sphere and success at work and, appropriately, intelligence and education. We refer to it as *Status*, but point out that the intellectual aspect should always be kept in mind.

The core of the fifth component is formed by withdrawn, sad, earnest and timid. There are also secondary loadings of in a good mood(-), merry(-), cheerful(-), whose main loadings are on the Agreeableness component, as well as insecure and nervous. The focus here is on mood, with negative affects carrying more weight than the lack of positive affects. There are also aspects of social withdrawal. As the core is defined by negative affects, we do not reverse the polarity in terms of desirability and refer to this component as *Negative Affects*.[8] We will occasionally recall that a high score on this component is undesirable.

The right-hand column of Table 2 shows that for most items the variance is explained to a very high degree by the five-dimensional space. Only for nine items is the proportion of unique variance more than 20 percent and the maximum is 33 percent (unpredictable).

We now have a five-dimensional space whose coordinates are orthogonal, and the five dimensions cover very different aspects of personality, such as dispositions, moods, abilities, social evaluations, social effects and physiognomic characteristics. It should be noted that our labels Agreeableness, Attractiveness, Masculinity, Status and Negative Affects in turn capture different aspects with different weightings and that the content of the components is much broader than can be expressed by a simple label.

In this space, which we refer to as *personality impressions space* or *personality space* for short, the stimulus persons can be located by means of their factor scores. This is illustrated in Figure 1 using the example of the plane spanned by the Agreeableness and Attractiveness components. Each dot there represents a man. Figure 4 in the supplement shows the same plane in which the variables are plotted instead of the men.

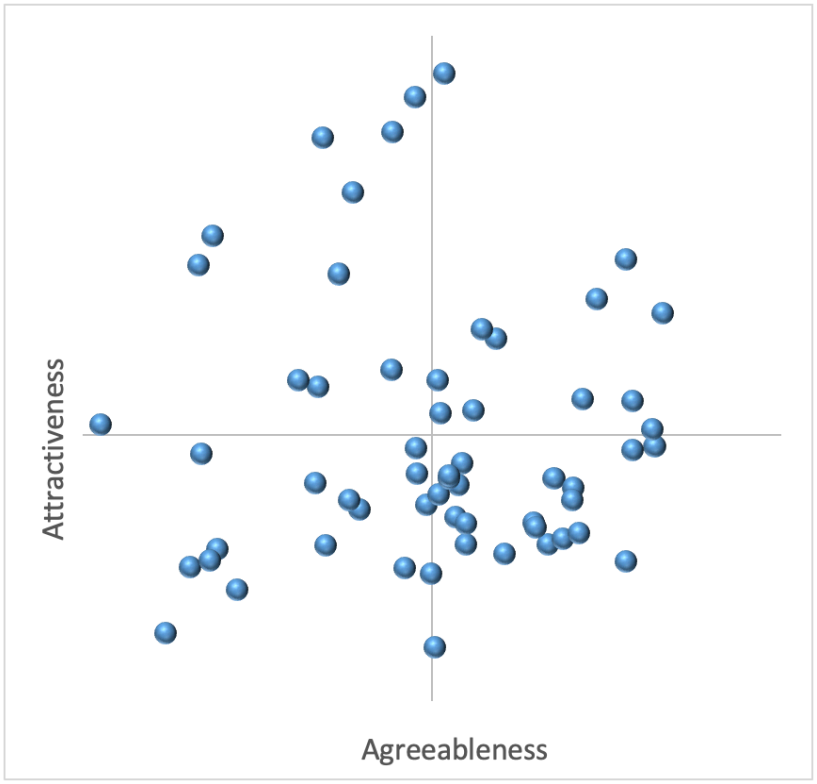


Figure 1. *Stimulus persons in the plane Agreeableness / Attractiveness.*

We will differentiate the space in more detail in a moment. Here we would like to note an important point. Our further analyses are based on the factor scores. These are standardized to a mean of 0 and a standard deviation of 1 and must not be confused with the absolute scores on the original rating scales. This is particularly important for Attractiveness. In Figure 1, twenty-three men are above the horizontal axis that represents the mean of the factor scores. On the original rating scales, however, the men score extraordinarily poorly. Averaged over the items sexy, good looking, erotic, seductive and successful with women, only eight men are above the middle of the scale. When we say 'This group is more attractive than that group' below, it would be more appropriate to say 'This group is less unattractive than that group' or 'That group is even less attractive than this group'. For the sake of convenience, however, we will use the positive mode of speaking.

Before broadening the field of view, we want to briefly address the exploratory factor analysis. As can be seen in Table 3, the first four factors match almost perfectly with the components of the principal component analysis and there is only a notable difference in the fifth factor. Here, in the explorative factor analysis, the axis is rotated so that the mood items are represented more appropriately.

doi.org/10.32388/6IDC9O

External variables in the personality space

With the principal component analysis, we have obtained a space with five independent dimensions that can be interpreted in a highly plausible way and covers a wide spectrum of personality impressions. We now broaden the scope and ask how the external characteristics relate to this space.

First, we look at the simple correlations between the external characteristics on the one hand and the personality factors on the other. In addition, the right-hand half of Table 5 shows the partial correlations after controlling for Age. Correlations that do not pass the 5 percent threshold are omitted.

	Simple correlation						**Controlled for Age**				
	Agr.	**Attr.**	**Masc.**	**Stat.**	**NAff.**		**Agr.**	**Attr.**	**Masc.**	**Stat.**	**NAff.**
Age		-.57	.54				---	---	---	---	---
Smile	.56	-.27			-.58		.53		-.32		-.59
Height	-.43	.33	.32				-.39		.61		
Weight			.35						.30		
BMI		-.39						-.26			
Hair Loss		-.38								-.28	
Hair Color			-.31								
Beard				-.43						-.42	
Glasses			-.36	.52					-.42	.53	
Tie		-.29		.27							
Tilt absolute				.38						.36	
Tilt											
Face Side											
Background											

Table 5. *External variables and personality factors. Left: simple correlations. Right: controlled for Age.*

We look at the table from the simple and clear to the more complex. The last three rows are the simplest. Tilt, Face Side and Background do not correlate with any personality dimension. Negative Affects have a single correlate under both conditions. The greater the degree of Smile, the lower the Negative Affects. Agreeableness correlates positively with Smile and negatively with Height in both conditions. Status has four significant correlates in each case. Without taking Age into account, Glasses, Tilt absolute and Tie have a positive effect, a Beard a negative one. After controlling for Age, the correlation with Tie is no longer significant, but the negative correlation with Hair Loss passes the 5 percent threshold. Masculinity has a remarkable positive correlation with Age (.54). Under both conditions, Masculinity correlates positively with Height and Weight and negatively with Glasses. The negative correlation with Hair Color is only significant as long as Age is not taken into account, while on the other hand Smile shows a significant negative effect only after controlling for Age. For Attractiveness, the situation appears quite different. As to be expected, there is a substantial negative correlation with Age (-.57). In addition, there are five further significant correlations. The relationship with Height is positive, and it is negative with BMI, Hair Loss, Tie and Smile. However, after controlling for Age, only the negative correlation with BMI remains.

Overall, controlling for Age has no effect at all on Negative Affects and Agreeableness and causes only minor shifts in Status and Masculinity. This is all the more true when comparing the size of the correlation coefficients. The changes are small, only the correlation between Height and Masculinity increases from.32 to.61. Correlations that are no longer significant after controlling for Age were already only barely above the 5 percent threshold, and correlations that are now significant were barely below it before. This also applies to Attractiveness. Only the correlation with Hair Loss (-.38) had passed the 1 percent hurdle. This means that the disappearance of the significant correlations looks more dramatic for Attractiveness than it actually is.

Overall, the relationships between external variables and personality factors are not moderated by Age at all or only to a limited extent. However, another effect should be noted at this point. As mentioned, Age correlates negatively with Attractiveness (-.57) and positively with Masculinity (.54). As both were extracted in a principal component analysis with varimax rotation, they are orthogonal. However, after controlling for Age, they correlate at.45. This shows that Masculinity can be considered a facet of Attractiveness and the orthogonality of the two principal components is a consequence of the method.

So far, we have only looked at pairwise correlations. Now we are broadening the view. More precisely, we ask whether the personality space can be mapped to the external characteristics and how well this mapping corresponds to the actual scores. To this end, we use multiple regression.

Using the example of the age estimates, this means that we are looking for that linear combination of the factor scores of the five components which has the highest correlation with the scores of our external variable Age. Here, the multiple regression leads to the equation

Age* = 42.746 + 1.899·Agreeableness - 4.462·Attractiveness + 4.226·Masculinity + 1.695·Status + 0.061·Negative Affects

The multiple correlation R, i.e. the product-moment-correlation between the estimate Age* and the given value of Age, is.853. Thus, 72.7 percent of the Age variance is explained by the personality space, adjusted 70.2 percent. Attractiveness and Masculinity have the highest weight, Agreeableness and Status also play a role, but Negative Affects is of no importance. Conversely, it can also be said that Age has a very close relationship with the personality space, particularly with Attractiveness and Masculinity, but also with Agreeableness and Status, but not with Negative Affects.

Table 6 shows the results for all external variables. The left part shows the parameters of the overall test, the right part shows the significance level of the individual components. Positive correlations are marked in green, negative correlations in red. It should be noted that the right-hand section of Table 6 is not to be understood as a customary correlation table. If the personality dimensions were considered as criteria and the external variables as predictors, there would be some shifts. Thus, the table should be interpreted by rows.

	R	R^2	Adj. R^2	p	Agree	Attr	Masc	Stat	Aff
Smile	.860	.739	.715	< .001	***	***	*		***
Age	.853	.727	.702	< .001	**	***	***	**	
Glasses	.670	.449	.397	< .001		*	***	***	
Height	.631	.398	.341	< .001	***	**	**		
Tilt absolute	.519	.269	.200	.004			*	**	*
Beard	.494	.244	.173	.009				***	
BMI	.480	.231	.158	.014		**			
Hair Loss	.474	.225	.152	.016		**			
Weight	.439	.193	.117	.040			**		
Tie	.403	.162	.083	.086		*		*	
Hair Color	.396	.157	.077	.099			*		
Background	.309	.096	.010	.360					
Tilt	.283	.080	-.007	.475					
Face Side	.263	.069	-.019	.564					
					* p < .05 ** p < .01 *** p < .001				

Table 6. *Multiple regression of external variables on principal components.*

doi.org/10.32388/6IDC9O

The external variables are represented very differently in the psychological space. For Smile, Age, Glasses and Height, the overall test is p <.001. For Tilt absolute and Beard, the multiple regression is significant at the 1 percent level, for BMI, Hair Loss and Weight at the 5 percent level. Tie and Hair Color only show a tendency in the overall test. Background, Tilt and Face Side cannot be reconstructed by a linear combination of the five components of this space. We will not consider these three variables any further.

By far the most important external variables are Smile and Age. Here, even the adjusted proportion of variance explained is more than 70 percent. Smile correlates at the 0.1 percent level with Agreeableness, Attractiveness(-) and Negative Affects(-) and at the 5 percent level with Masculinity(-). Age correlates with Attractiveness(-) and Masculinity at the 0.1 percent level and with Agreeableness and Status at the 1 percent level.

Glasses and Height are also significant at the 0.1 percent level, their adjusted variance explanation is 39.7 and 34.1 percent respectively. Glasses correlates with Masculinity(-) and Status at the 0.1 percent level and with Attractiveness(-) at the 5 percent level. Height correlates with Agreeableness(-) at the 0.1 percent level and with Attractiveness and Masculinity at the 1 percent level.

Cross validation

Multiple regression is a very powerful tool and the question arises to what extent the results can be generalized. To test this, we carried out a cross-validation. The men with an even identification number served as the calibration sample, while the men with an odd identification number served as the validation sample. In the first step, a multiple regression was carried out with the calibration sample. In the second step, the parameters obtained were applied to the validation sample and this estimate was correlated with the actual scores.

Table 7 shows the results of the successful replications. The left part shows the multiple regression coefficients and the significance level for the calibration sample, while the right part shows the values for the validation sample.

	Calibration sample		Validation sample	
	R	p	r	p
Age	.86	<.001	.86	<.001
Smile	.95	<.001	.72	<.001
Glasses	.62	.036	.64	<.001
Height	.71	.004	.56	.001

Table 7. *Cross-validation.*

Cross-validation was successful for Age, Smile, Glasses and Height. Based on the parameters of the calibration sample alone, there is a significant prediction at the 0.1 percent level in the validation sample and for Age and Glasses the predictive power in the validation sample is as high as in the calibration sample.

Cross-validation was not successful for Tilt absolute, Beard, BMI, Hair Loss, Weight, Tie and Hair Color. Here, the multiple correlation in the overall sample is already lower, so that a successful cross-validation could hardly be expected for the halved samples. As these variables correlate with at least one component of the personality space, they will continue to be taken into account in the following.

Typological perspective

Now we change perspective and turn to the typological approach. The starting point is the simple assumption that some people are more similar to each other than to others and that groups of men can be identified who are similar to each other in a number of ways and who differ markedly from other groups. To construct such groups, we use both the dimensions of the personality space and the external variables.

Based on the standardized variables, we conducted a hierarchical cluster analysis using a Euclidean distance measure and the clustering method Ward.D2. The dendrogram is shown in Figure 2.

doi.org/10.32388/6IDC9O

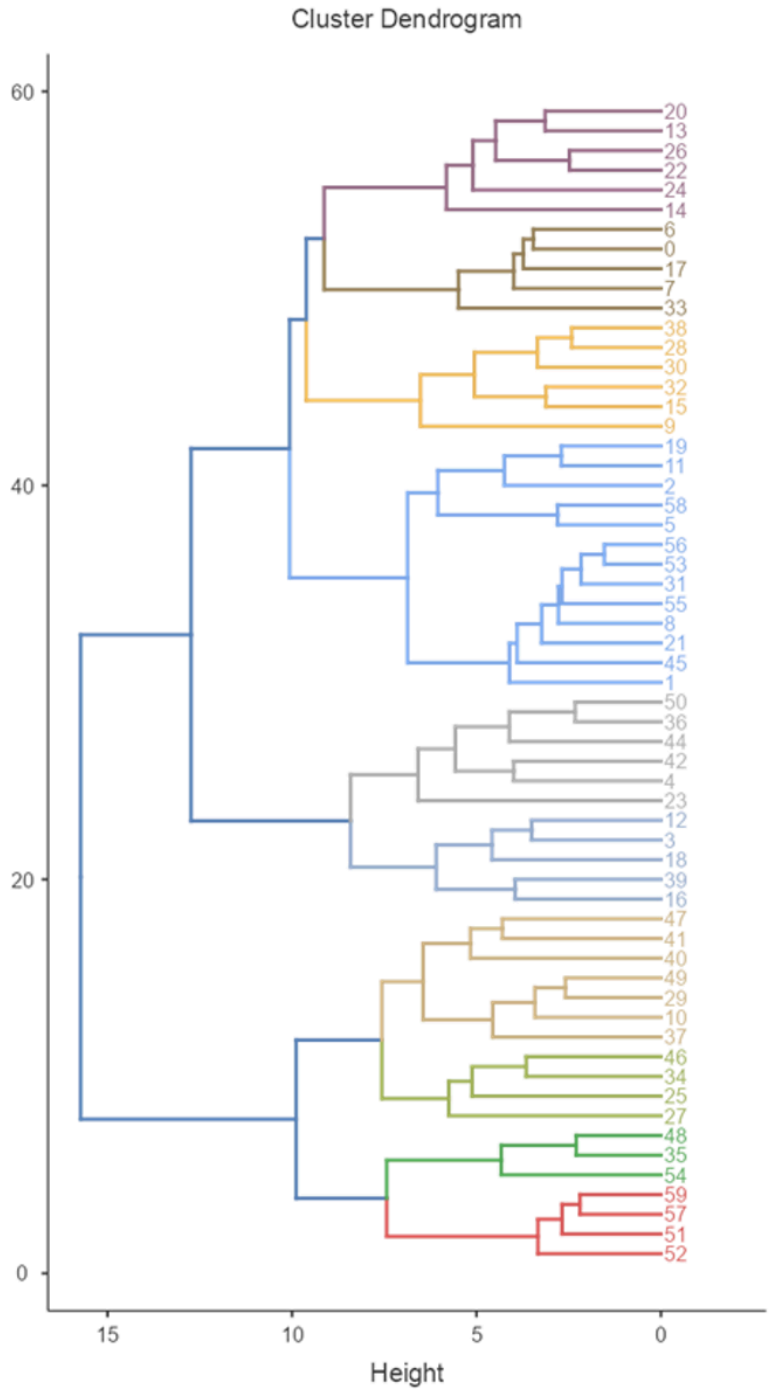
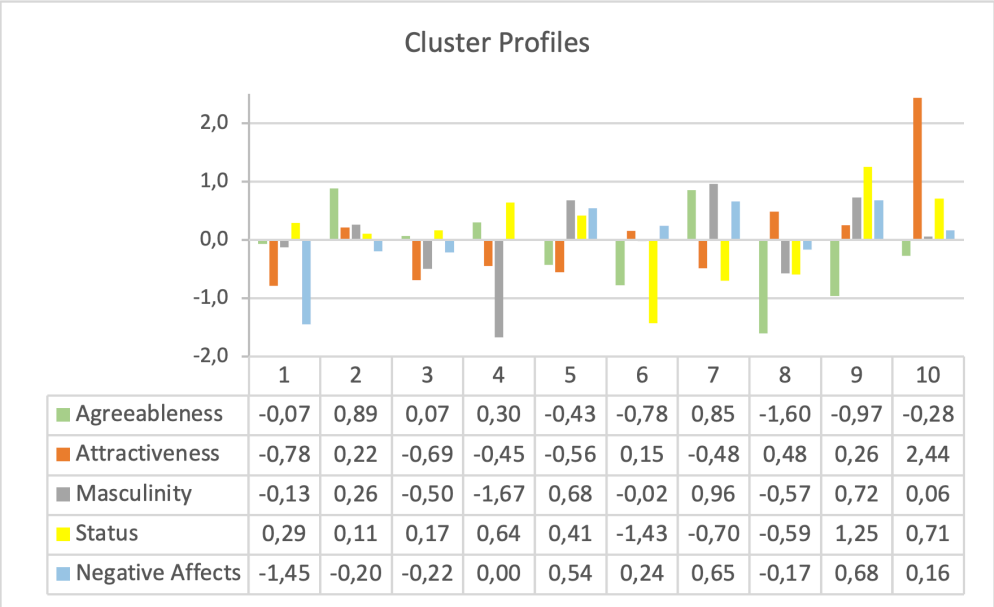


Figure 2. *Hierarchical cluster analysis. Dendrogram.*

For a coarse-grained analysis, one would probably opt for a 3-cluster solution. However, we are interested in a fine-grained analysis and have opted for the 10-cluster solution. The main reason for this can be seen in

doi.org/10.32388/6IDC9O

Figure 3. This shows the average factor scores of the components of our personality space broken down by the clusters.



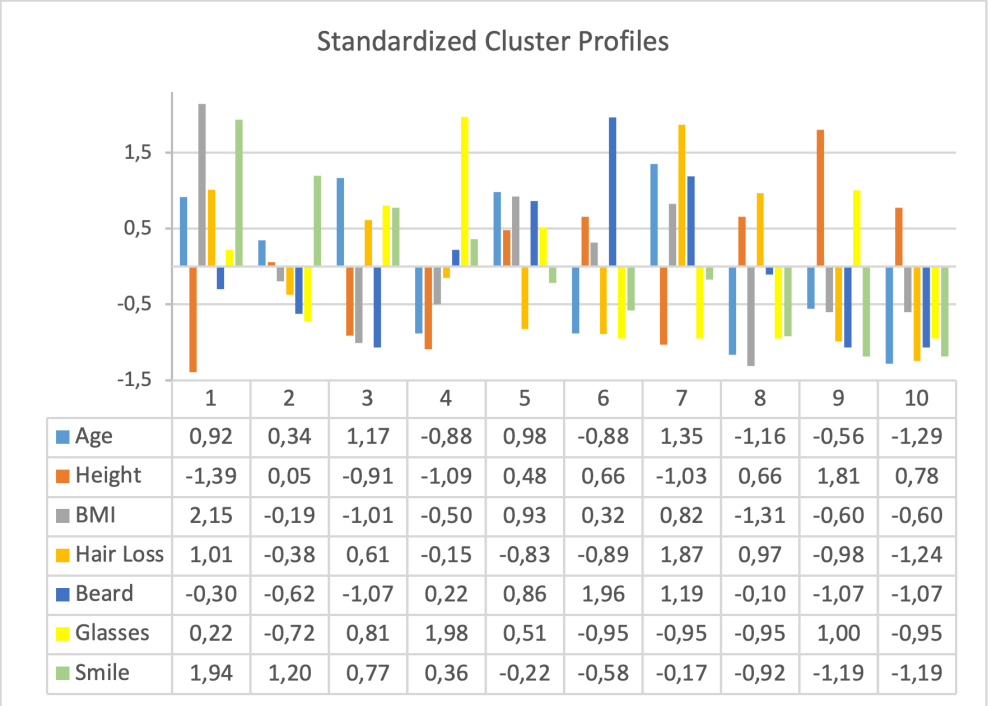
	1	2	3	4	5	6	7	8	9	10
Agreeableness	-0,07	0,89	0,07	0,30	-0,43	-0,78	0,85	-1,60	-0,97	-0,28
Attractiveness	-0,78	0,22	-0,69	-0,45	-0,56	0,15	-0,48	0,48	0,26	2,44
Masculinity	-0,13	0,26	-0,50	-1,67	0,68	-0,02	0,96	-0,57	0,72	0,06
Status	0,29	0,11	0,17	0,64	0,41	-1,43	-0,70	-0,59	1,25	0,71
Negative Affects	-1,45	-0,20	-0,22	0,00	0,54	0,24	0,65	-0,17	0,68	0,16

Figure 3. *Cluster profiles. Factor scores of the personality factors.*

Cluster 10 on the right-hand side stands out due to the unusually high orange bar that indicates Attractiveness. No other variable shows such a large gap to the other clusters as here.

Figure 4 shows the standardized profiles of the external variables that are of particular interest.

doi.org/10.32388/6IDC9O



	1	2	3	4	5	6	7	8	9	10
Age	0,92	0,34	1,17	-0,88	0,98	-0,88	1,35	-1,16	-0,56	-1,29
Height	-1,39	0,05	-0,91	-1,09	0,48	0,66	-1,03	0,66	1,81	0,78
BMI	2,15	-0,19	-1,01	-0,50	0,93	0,32	0,82	-1,31	-0,60	-0,60
Hair Loss	1,01	-0,38	0,61	-0,15	-0,83	-0,89	1,87	0,97	-0,98	-1,24
Beard	-0,30	-0,62	-1,07	0,22	0,86	1,96	1,19	-0,10	-1,07	-1,07
Glasses	0,22	-0,72	0,81	1,98	0,51	-0,95	-0,95	-0,95	1,00	-0,95
Smile	1,94	1,20	0,77	0,36	-0,22	-0,58	-0,17	-0,92	-1,19	-1,19

Figure 4. *Standardized cluster profiles. External variables.*

As in Figure 3, we can also see here that the cluster profiles differ substantially from one another.

In the following, we take a closer look at the clusters. In order to keep things transparent, we only consider the five dimensions of the personality space in the graphical representation. However, we will also mention the relevant external variables in the description of the clusters.

Figure 5 shows the profiles of clusters 1, 3, 8 and 10.

doi.org/10.32388/6IDC9O

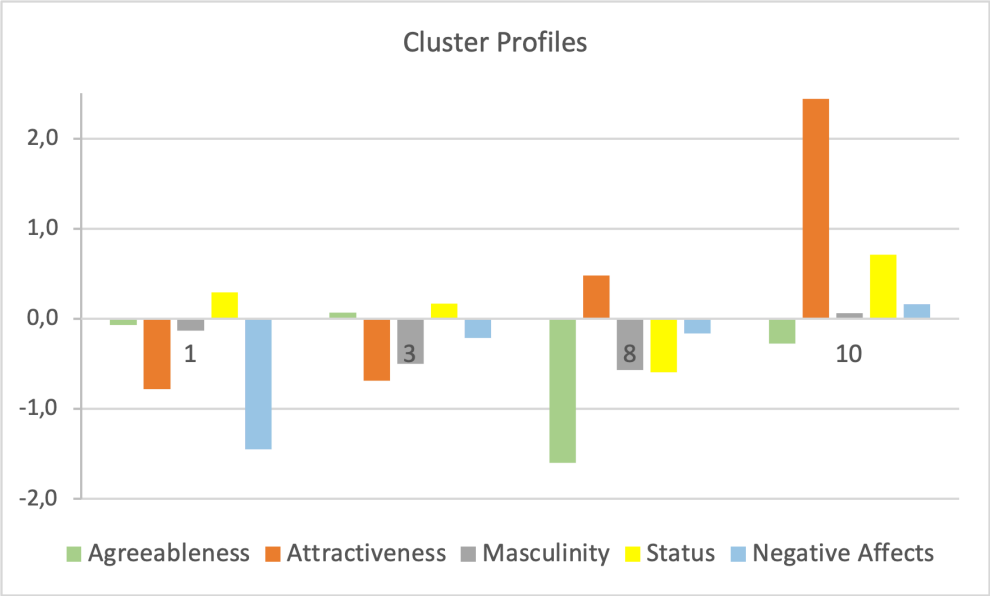


Figure 5. *Cluster profiles. Clusters 1, 3, 8, 10.*

With these four clusters, our attention is on Attractiveness, which is represented by the orange bars. Two important points should be recalled here. Firstly, the focus of this component is on the sexual aspect. Secondly, we are only looking at the relative standing of the clusters and should remember that the men scored extremely poorly on the original attractiveness items – the attractiveness malus for men.

Cluster 10 comprises the most attractive men by a huge margin. This becomes clear when compared with cluster 8, which is the second most attractive group. Cluster 1 is the least attractive and cluster 3 the second least attractive. Figure 5 thus illustrates the extremes on the Attractiveness component.

The attractive men in cluster 10 achieve the second-highest rank for Status and are in the middle range on the other three dimensions. They form the youngest group and rank second for Height, they show no sign of Hair Loss and wear neither a Beard nor Glasses and they show no sign of Smiling.

Cluster 8 ranks second in Attractiveness, but at the same time these men score by far the worst in Agreeableness, the second worst in Masculinity and the third worst in Status. Here, attractiveness is paired with an otherwise very poor rating. These men make up the second youngest group, but they show the third highest degree of Hair Loss. They have the lowest BMI and none of them wears Glasses and only one of them has a light stubble beard.

Cluster 1 and Cluster 3 rank the lowest and second lowest on Attractiveness. They also have the lowest and second-lowest scores on Negative Affects. However, since Negative Affects are poled differently, this means

that although these men are very unattractive, they are the best in terms of mood. This rating is certainly mediated by the fact that Cluster 1 ranks first in Smile and Cluster 3 ranks third. Cluster 1 is the shortest and most corpulent and shows the second highest degree of Hair Loss. Cluster 3, on the other hand, includes the slimmest men and they are the second youngest group.

Figure 6 shows the profiles of clusters 2, 6, 7 and 9.

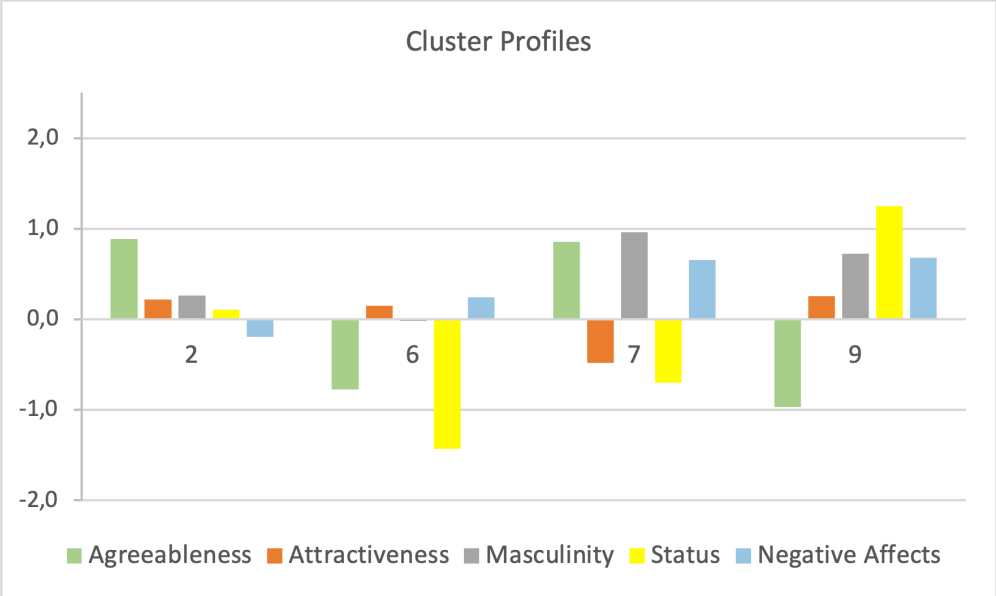


Figure 6. *Cluster profiles. Clusters 2, 6, 7, 9.*

In this quartet, our focus is on Agreeableness.

Cluster 2 achieves the top position on Agreeableness and second place on Smiling. For all other variables, it ranks in the middle, making it the most colorless. However, this is probably mainly due to the fact that this cluster comprises thirteen men and is therefore the most heterogeneous.

Cluster 7 is in second place on Agreeableness, ranking highest on Masculinity, second lowest on Status and second highest on Negative Affects. These men are the oldest, they show the most severe Hair Loss, they are in second place among beard wearers and they do not wear glasses. In contrast to the colorless Cluster 2, very high Agreeableness is associated with a distinctive profile here.

Cluster 9 ranks second to last for Agreeableness, along with the strongest expression of Negative Affects and last place on Smile. On the other hand, this cluster ranks first on Status and second on Masculinity. These men are estimated to be the tallest. They have the second lowest degree of Hair Loss and they do not wear facial hair. Two of the three men wear glasses and this trio shows the strongest head Tilt.

Cluster 6 ranks third to last on Agreeableness. The outstanding feature is the lowest rating on Status by a wide margin. This group contains the highest proportion of beard wearers, but no spectacle wearers.

We have not shown clusters 4 and 5. Cluster 4 stands out due to its exceptionally low Masculinity. This group ranks first for Glasses and second to last for Height. Cluster 5, like cluster 2, is quite colorless. It is in second place on Weight, BMI and Tilt, otherwise it ranks in the middle.

Finally, one point should be highlighted that is not apparent so far, but which catches the eye in Figure 7. The figure shows the mean values and confidence intervals of Smile, with the clusters arranged in the order in which they were determined by the hierarchical cluster analysis.

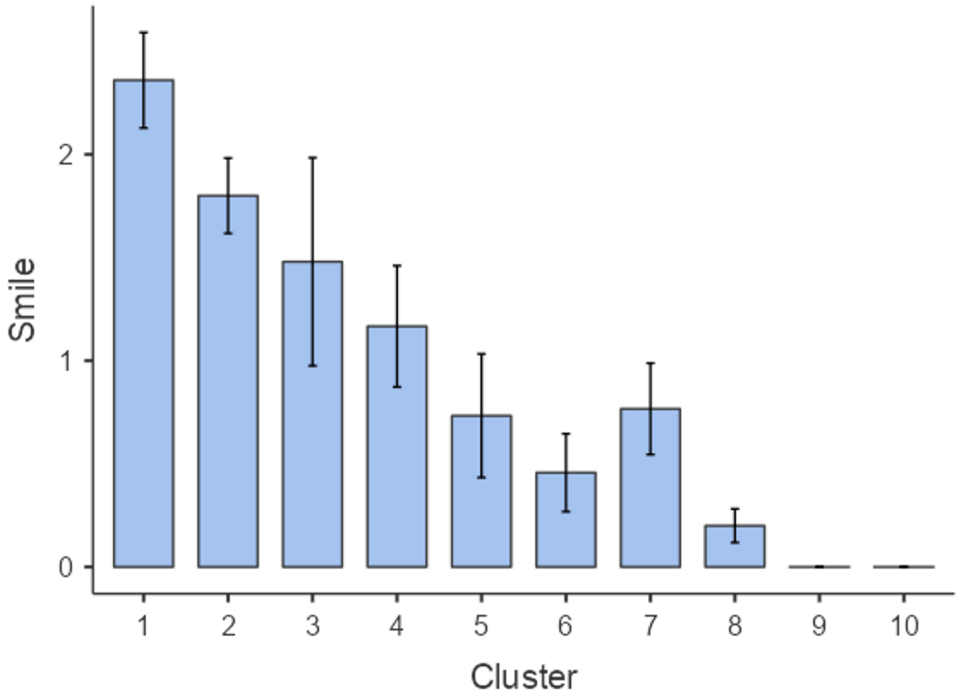


Figure 7. *Smile and clusters, arranged according to the hierarchical cluster analysis.*

The rankings of clusters 1 to 10 for Smile are 1, 2, 3, 4, 6, 7, 5, 8, 9.5 and 9.5. This shows that the degree of smiling played a predominant role in cluster formation. It can also be seen that there are fundamental differences between the clusters with regard to smiling, with only a couple of pairs showing overlapping confidence intervals. A significant, but not as close, relationship can be seen for Height, Attractiveness and Negative Affects (see Table 3 in the supplement).

doi.org/10.32388/6IDC9O

Individuals, Clusters, Dimensions

Finally, we bring the three building blocks of our analysis – individuals, clusters and dimensions – into a common framework. Figure 8 shows the individuals and the clusters in the plane spanned by the components Agreeableness and Attractiveness, which we have already considered in Figure 1. Now, however, the clusters are marked in color.

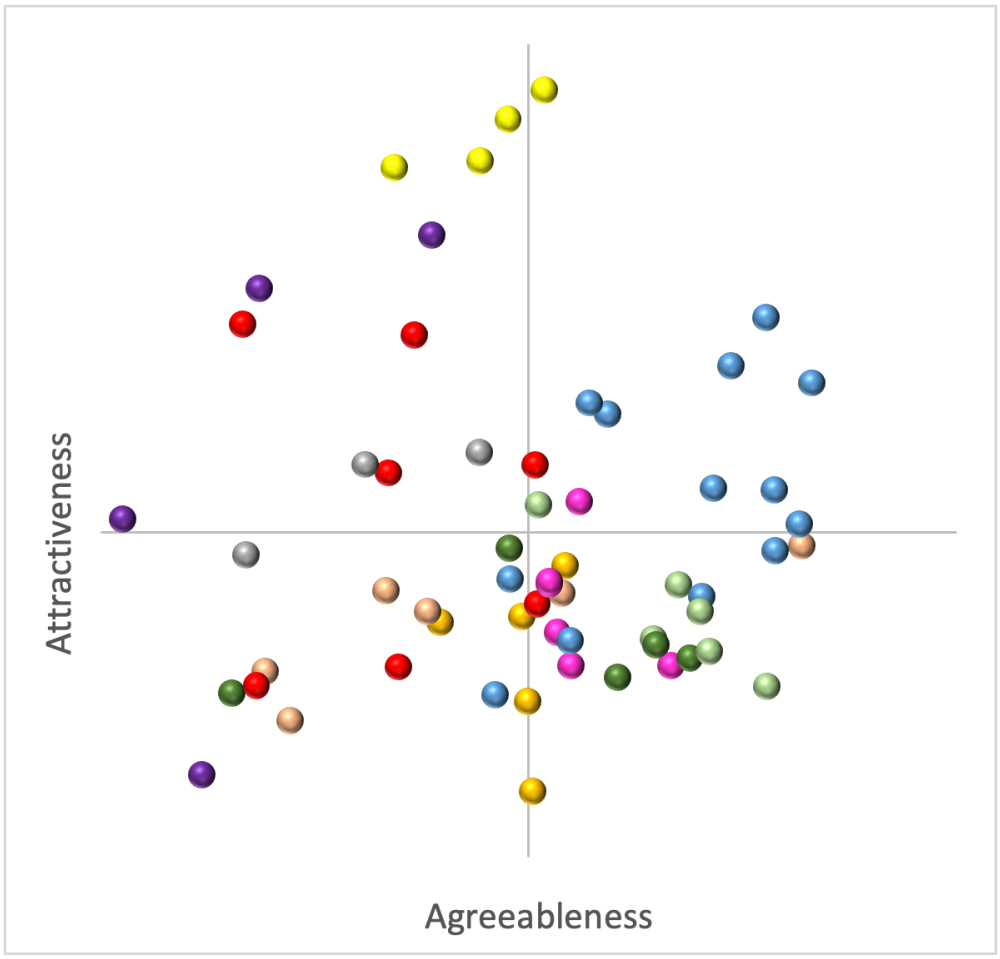


Figure 8. *Individuals and clusters in the plane Agreeableness / Attractiveness.*

A first thing to notice is that the clusters do not occupy strictly separated areas. Only the four most attractive men, who form the cluster marked in yellow, occupy an area that is not reached by anyone else. Secondly, it can be seen that the clusters overlap with some others on the one hand, but are disjunct to others on the other hand. The number of disjunct pairs is greater than the number of overlapping pairs. Thirdly, the clusters differ in their compactness. While the individuals in some clusters are close together, other clusters occupy a large area. The latter means that the members of the group are not very similar in terms of Agreeableness and/or

doi.org/10.32388/6IDC9O

Attractiveness. Fourthly, most clusters are only found in two quadrants, only the blue one in three and the red one in all four quadrants, but the latter only by a very narrow margin.

We now look at the Masculinity / Status plane in Figure 9.

Here we see a picture that is very similar on the one hand, but very different on the other. The clusters are still of varying compactness and there are overlaps and disjunct cluster pairs. Some clusters that overlap in Figure 8 continue to do so. Another point is decisive: Some clusters that overlap in the Agreeableness / Attractiveness plane are now far apart, and others that were far apart now overlap. For example, the red and gray clusters have drifted apart, as have the light green and pink clusters. On the other hand, the yellow and gold clusters, which are the extremes on the Attractiveness axis, are now barely distinguishable from each other. In the same way, the extremes on the axis of Agreeableness (dark purple and blue) have moved closer together. Figure 10 only shows the cluster centers in the two planes. Here, the radical shifts can be seen much more clearly.

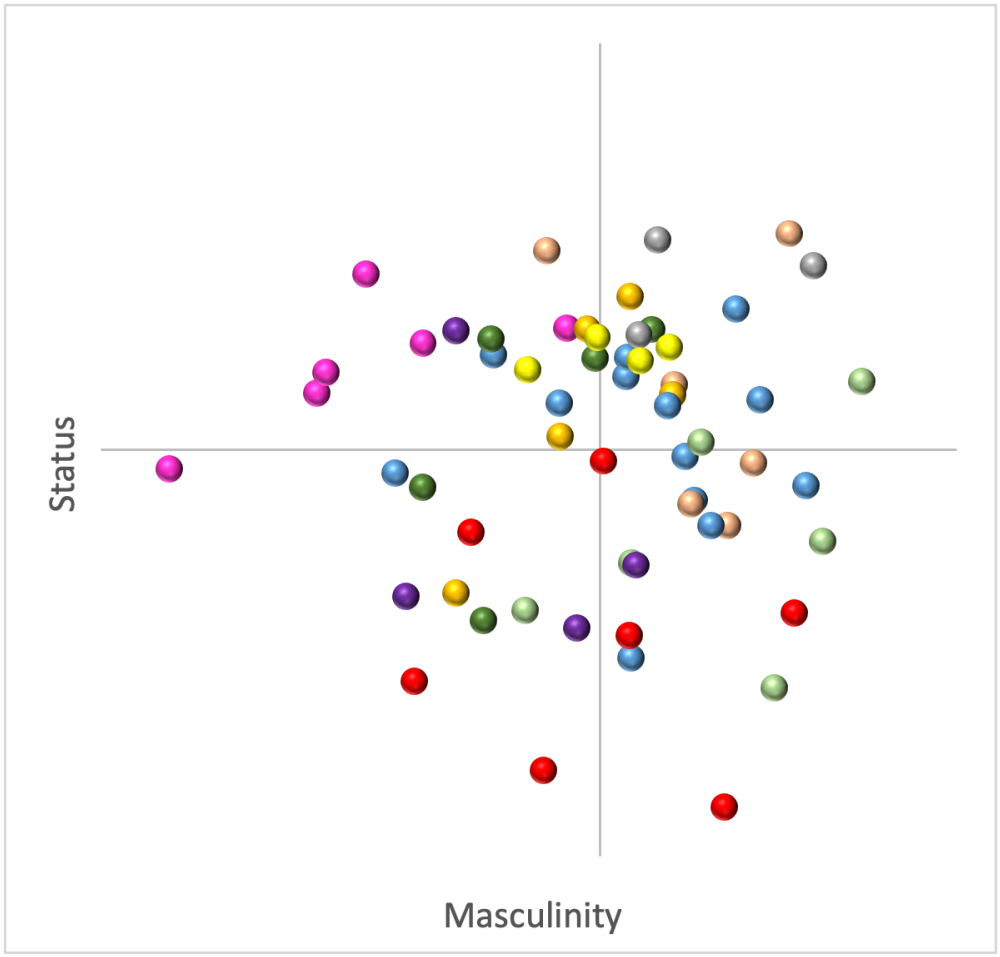


Figure 9. *Individuals and clusters in the plane Masculinity / Status.*

doi.org/10.32388/6IDC9O

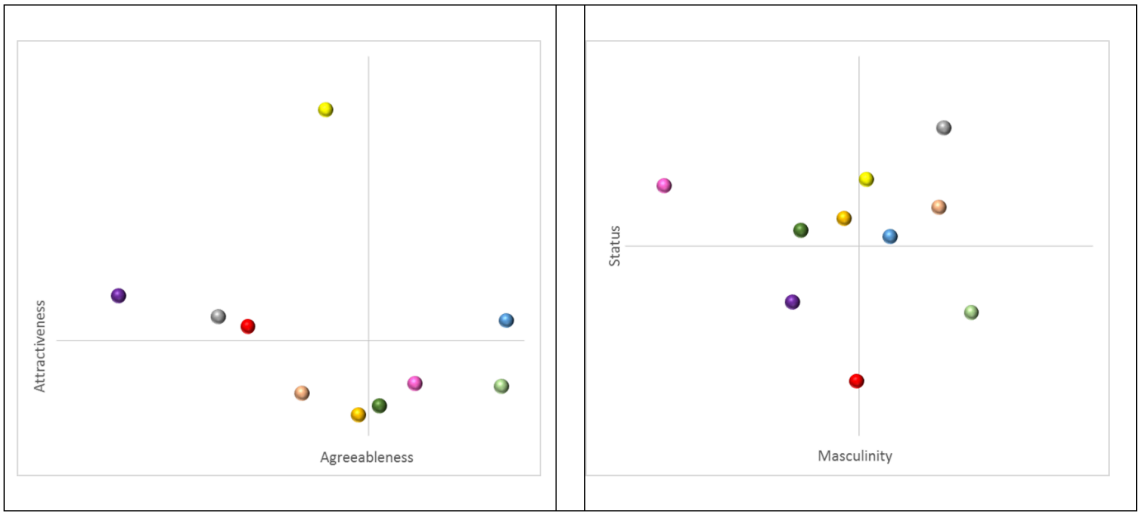


Figure 10. *Cluster centers in the planes Agreeableness / Attractiveness and Masculinity / Status.*

If one also adds the fifth component, Negative Affects, then it is evident that the clusters in the full personality space form well-defined clouds of dots. One just has to look at the overall picture from the right angle and not rely on a single plane. In addition, there are the differences in the external characteristics, which are also reflected in the personality space.

The comparison of figures 8, 9 and 10 should have made it clear how fruitful the combination of the dimensional and the typological perspective is. Both together provide insights that neither can offer on its own.

Discussion

The subject of this paper is social judgment of men based on facial photographs. The restriction to men stems from the fact that our dataset is based on a study on effects of genetically induced hair loss in men (Henss, 2024a). This is important to emphasize because the personality domain we consider was steered in a certain direction by the initial research questions. Some aspects that interested us in connection with hair loss were brought into focus, while others were excluded. In the current study, hair loss is merely one variable among others. We have supplemented the data set with some additional variables and analyzed them from other perspectives.

Our basic topic is the structure of personality impressions and we examine this from two perspectives. The first is a dimensional representation of the relationships between the various features that we perceive when looking at another person's face. To this end, we used factor-analytical methods to construct a low-

doi.org/10.32388/6IDC9O

dimensional space and interpret its content. This perspective is daily business in research on face perception. In addition, we analyzed the data from a typological perspective, which is extremely rare in this branch of research. The aim here is to use cluster analytic methods to identify groups of people whose members are similar to each other in a bundle of characteristics and who are clearly different from other groups. We have not only shown that both perspectives provide valuable insights, but that they can also be combined in a fruitful way. In this way, conventional research in particular is enriched by as yet largely untapped possibilities.

Before approaching the core questions, we look at the reliability of our baseline data; more specifically, we look at the degree of consensus in judging other people's faces as a function of the characteristic under consideration. Here, two test arrangements must be distinguished. The main experiment took place as a single stimulus assessment. Each subject judged a single randomly selected face with regard to 36 personality-descriptive characteristics and estimated age, body height and weight. In addition, a number of external characteristics were determined through serial stimulus assessment, i.e. each subject assessed all of the photographs.

In the case of serial stimulus assessments, agreement is already very high at the level of individual judges, the agreement of the individuals with the average judgment of the others is necessarily even higher and the reliability of the group standard is almost perfect. On the one hand, this is due to the fact that we are dealing here with features that are very easy to recognize in photos, namely the extent of hair loss, hair color, degree of smile, head rotation, head tilt and brightness of the background. Secondly, the photos provide a common frame of reference. Thirdly, the agreement between judges is determined by correlations, and differences in the absolute level do not play a role here. These three conditions are not fulfilled in single stimulus judgments and, as a result, the consensus is much lower. This is particularly noticeable at the level of individual judges. At first glance, the consensus here may appear to be extremely low. We will come back to this in a moment. First, we note that our analyses are based on the level of the stimulus persons, i.e. for each characteristic, we consider the average of all judges who rated the photo in question. Even in the case of single stimulus assessments, these group standards have sufficient, usually high or even very high reliability. Only a few items have insufficient reliability. However, since we do not look at the individual items, but at personality factors, which have a higher reliability, our analyses are throughout based on reliable to extremely reliable data.

Of particular interest are the differences in the psychological characteristics, which can only be assessed subjectively. The highest consensus is found for positive affects. Next in line is attractiveness. Quite contrary to the adage „Beauty is in the eye of the beholder", attractiveness is among the psychological characteristics

that are assessed with the greatest unanimity. This is followed by the professional sphere, then indicators of masculinity, then the area of social agreeableness. Clearly lagging behind are the closely related areas of emotional lability and negative affects. It is precisely this pattern that we have repeatedly observed in our studies (Henss, 1998a) and it can also be found in other studies on face perception.[9]

Of particular note is the extreme contrast between positive and negative affects. Contrary to what the name suggests, positive and negative affects are not the opposite poles of a common dimension, but two almost orthogonal factors (Watson, Clark and Tellegen, 1988; Henss, 2024c).[10] Positive mood is rated with the highest concordance. The explanation is obvious: The items merry, cheerful and in a good mood correlate very highly with the degree of smiling (.85,.84,.82) and this is so clearly recognizable in photos that it is rated with an extraordinarily high concordance. Negative affects, on the other hand, are difficult to recognize in photos. Dynamic mimic cues or vocal cues are apparently required here (Borkenau and Liebler, 1992). There is also another banal reason for this. Our photos were taken from a model catalog. Both the models themselves and the model agency will have attached great importance to conveying a positive impression. The same applies to people who volunteer to be photographed for research purposes; and in many studies, there is an explicit emphasis on a neutral facial expression. It is therefore no wonder that negative emotions are difficult to see in social psychological research on face perception. The situation is quite different in the clinical field or in emotion research.

Although this is not our core topic, we would like to extend our consideration of concordance among judges. Researchers in this field are typically not interested in the evaluation by individual judges, but in the average evaluation by the group. Accordingly, they usually report Cronbach's α or the intra-class correlation ICC(., k) as the reliability measure, depending on the test design. This is perfectly okay. However, since these measures depend on the number of judges, they do not allow a direct comparison between different studies. In order to compare different studies, the values must be converted to the individual level, i.e. to the average inter-individual correlation r(i, j) or ICC(.,1). Unfortunately, this information is rarely provided in the literature.

In order to give an impression of how the degree of inter-individual consensus is to be interpreted, we want to look at single stimulus assessments, which provide markedly lower values than serial stimulus assessments. Our ICC(1,1) values for the subjective psychological traits range from 0.03 to 0.34, with a median of 0.15. The ICC(1,1) is determined in a one-way analysis of variance without repeated measures. In this design, the coefficient $f = \mathrm{sqrt}(\eta^2 / (1 - \eta^2))$ is often used as a measure of effect size. Substituting ICC(1,1) for η^2 results in a range of 0.18 to 0.72 with a median of 0.42. Since ICC(1,1) is lower than η^2, these values are an underestimate. According to a suggestion by Cohen (1988, 1992), an f value of 0.10 is considered a small effect, an f of 0.25 a medium effect and an f of 0.40 a large effect. Accordingly, our minimum would correspond to a small to

medium effect size, the median to a large effect size and our maximum to an extraordinarily large effect size. Measured against the effect sizes that are common in psychological research, we are therefore dealing with high to very high agreement between judges. Hence, one should not be deceived by the seemingly small numbers.

For reasons of comparability, one should always report the agreement at the individual level in addition to the reliability of the group standard. The supplement contains a conversion table from the individual to the group level based on the Spearman-Brown formula. There it becomes clear that for many characteristics one can obtain very reliable ratings at the group level with remarkably small samples. As a rule of thumb, a reliable to very reliable group standard can be obtained for most characteristics with two dozen judges, even in single stimulus assessments where there is no common context. Even fewer are needed for serial stimulus assessments.

Valence-Dominance model

We now come to our core topic of the structure of personality impressions and first consider the dimensional perspective. As in the introduction, we take the Valence-Dominance model as our starting point. We have no doubt that it is of paramount importance where we locate a face on the Badness – Goodness dimension. Nor do we doubt the importance of the Dominance or Power dimension. Nevertheless, the model is hopelessly undercomplex for our research questions.

The Valence-Dominance model is a necessary consequence of the underlying item selection. To understand this, three points must be considered. First, a factor is a linear combination of variables that have something in common. Second, each personality trait has a more or less strong content reference as well as a more or less strong evaluative aspect. Third, Oosterhof and Todorov only considered a very small number of items and deliberately selected them to represent different areas of personality. As a result, the evaluative aspect remained the only commonality and inevitably an all-powerful evaluation factor had to emerge that explains the lion's share of the variance. If, on the other hand, a larger number of items are employed, with different areas each represented by several items, the content-related reference dominates over the evaluative aspect. Instead of one evaluation factor, there are then several factors that are defined by their common content, and none of them explains the lion's share of the variance.

If one is interested in the content structure of personality impressions, the Oosterhof-Todorov strategy is counter-indicated. Here, the impressive multiplicity of personality impressions is glossed over by a black-and-white painting. To avoid misunderstandings, we emphasize once again that each characteristic has a more or less strong evaluative aspect. The only common denominator across all characteristics is the evaluative

component, even if it is very small in some cases. Thus, it is neither coincidence nor arbitrariness that evaluation plays a prominent role in many models. These considerations make it clear that there is no such thing as the one and only true personality structure. Unfortunately, this simple fact is often forgotten.

For our original question about social perception of genetically caused hair loss in men, the Valence-Dominance model – or more precisely: the Valence dimension – would be of no interest as we know well enough that the overall effect of baldness is negative. However, we also know that the effects of hair loss depend on the personality trait in question (and also on the respective individual). Therefore, a differentiated approach was required from the outset. For our current research questions, a differentiated consideration of the personality area is virtually a necessary prerequisite. However, it should be emphasized here that it was not the initial aim to look at personality in its breadth. Our item selection was only aimed at a few areas that interested us in the context of our other studies. This means that although our data set enables a more differentiated picture than the 13 items of the Valence-Dominance model, several important aspects of personality are excluded in advance.

While the Valence dimension is too undifferentiated for our purposes, the Dominance or Power dimension is very important. For this, we want to take a closer look at the item dominant. As one would expect, it can be found on our Agreeableness factor. The loading of -.71 is remarkably high, but 12 items have an even higher loading. This means that dominance is not the core of this factor.[11] Nevertheless, the item dominant deserves special attention. The correlation with the 35 other personality-descriptive items is significant in 25 cases at the 0.1 percent level and in 2 cases at the 1 percent level. This means that the dominance aspect permeates wide parts of personality impressions. There is no correlation with earnest, withdrawn and sad as well as with intelligent, educated, successful in his job, career oriented and high occupational status. It is surprising that it is just the professional sphere and intelligence and education that are exempt. At the level of personality dimensions, dominant has a significant secondary loading on Masculinity (.52) and Attractiveness (.32). There are three significant correlations with our external characteristics. For Height (.58) and Smile (-.57), the correlation is significant at the 0.1 level and Weight (.26) narrowly passes the 5 percent threshold.

We had included the item dominant (and the related items aggressive, belligerent, dangerous, unpredictable) because it is of interest in connection with hair loss. While some authors suggest that a bald head is a signal of dominance and threat (Guthrie, 1976), others argue that a bald head is reminiscent of a baby face and is a signal of appeasement (Muscarella and Cunningham, 1996). In our sample, the correlation points in the direction of dominance, but it is not statistically significant (.24). Our answer to the controversy is Solomonic: It depends on the respective type. There are configurations in which a bald head has a placating effect, and

others in which it appears threatening, and still others in which neither is the case. It would be no problem to select faces so that one or the other side of this debate has the upper hand.

With regard to our current study, it should be stressed that dominance or power is a fundamental aspect of personality impressions. In any reasonably appropriate personality impressions space, the item dominant will be identifiable as a definite spatial direction, even if it does not necessarily fall exactly on one component in a principal component analysis.

Non-simplicistic factorial models of personality impressions

Todorov and Oh (2021) explicitly point out that their simplistic 2D model makes no claim to being complete and that a „different and a larger set of traits will result in a different dimensional solution“ (p. 214). Of particular interest is the sentence that immediately follows: „The only safe bet is that the first dimension will be about valence, something that we have known since the seminal work of Charles Osgood in the 1950s” (Todorov and Oh, 2021, p. 214). No, precisely that is not the case. As soon as a sufficient number of items are taken into account and – this is the crucial point – these are compiled in such a way that different areas are represented by several items each, there can be no question of an evaluation factor. Personality impressions are highly differentiated and not a crude black-and-white painting. It is the common content that dominates and not the purely evaluative aspect.

We have carried out numerous studies in our laboratory under a wide variety of conditions and have never seen anything resembling an evaluation factor. What’s more, the content of the respective factors was very similar and the pattern is in excellent agreement with the findings of differential psychology, which relies on self-assessments or acquaintance ratings. Notable differences only occurred when the item lists covered different areas from the outset, because we compiled them with regard to the respective research questions.

As mentioned in the introduction, the Big Three of Face Assessments are Attractiveness and the Big Two of the interpersonal domain, Extraversion/Mood and Agreeableness. Agreeableness and Attractiveness are also the two strongest factors in the present study. It was clear in advance that we would not find an Extraversion/Mood factor because extraversion was not represented by corresponding marker items. If, for example, we had added sociable, talkative, gregarious, adventurous and open, these would have combined with in a good mood, cheerful and merry to establish an Extraversion/Mood factor. Since extraversion items were missing, the positive mood items are found on Agreeableness. It should be noted at this point that in an exploratory factor analysis, in a good mood, cheerful and merry no longer load on Agreeableness, but together with sad, withdrawn, earnest and timid form a separate factor, and that Agreeableness no longer explains 29.8 of the variance, but only 22.3 percent.

One might argue that our principle component Agreeableness would be an evaluation factor after all. However, this would be a misnomer. The hallmark of this factor is not evaluation, but the content. This factor relates to the social domain and it would actually be more appropriate to call it Social Agreeableness, as we have occasionally done. As mentioned, we focused on the domain of dominance or antagonism. In the Valence-Dominance model, Valence and Dominance are orthogonal; in our model, Social Agreeableness and Antagonism are opposite poles of the same dimension. With the item selection, we have given particularly strong weight to this factor. Nevertheless, it only explains 29.8 percent of the variance and not, as Valence in the Valence-Dominance model, more than 60 percent. In addition, there is another important aspect. The negatively valenced traits timid, naive and babyface have a substantial positive secondary loading and the positively valenced traits masculine appearance and seductive have a substantial negative secondary loading. What's more, sexy, good looking, erotic, successful with women, mature face, pronounced male face, high occupational status and career oriented also have a negative sign, although they are undoubtedly desirable in men. All this makes it clear that Agreeableness is defined by the content and not by evaluation.

Our original focus was on the impact of male pattern baldness and, having already investigated this in other studies, our main interest was in Masculinity and Status/Intelligence. As we represented these with a sufficient number of items, we were able to identify both factors, as was to be expected.

Our fifth factor is essentially constituted by mood items, with negative emotions in the foreground and positive emotions having substantial secondary loadings. This factor is closely related to the Big Five factor Emotional Stability or Neuroticism. In many of our studies, we were only able to identify this factor in rudimentary form. We have discussed some of the reasons for this above in connection with the extraordinarily low level of consensus in judgments of negative affects.

At this point we would like to emphasize the following. Face perception is extraordinarily differentiated and there is no such thing as a single social judgment space. One only gets what one has put into it with the item selection. If only a dozen items are taken into account, there is hardly a chance of obtaining a higher-dimensional space. This is all the more true if, like Oosterhof and Todorov, one compiles the items in such a way that they cover very different aspects. In this case, one inevitably obtains an overpowering evaluation factor. If, on the other hand, a sufficient number of marker items are taken into account, it is an easy task to deliberately create a higher-dimensional space whose dimensions are defined in terms of content. These can be very broad factors, but also very specific narrow factors. Anyone interested in a differentiated view of personality impressions can utilize the treasure trove of insights from differential psychology. There one can find detailed factorial personality models and extensive lists of words or short descriptions that are suitable as markers for the various factors. Since facial impressions lead to a very similar structure as self-assessments or

acquaintance ratings, a dimensional representation can be constructed according to the modular principle as required.

The Attractiveness Stereotype „What is beautiful is good“

We began our research on face perception at the end of the 1980s with the question of consensus in attractiveness assessments (Henss, 1987) and subsequently developed a mate value theory based on Brunswik's lensmodel and evolutionary psychology, which focuses on physical attractiveness and its various facets (Henss, 1992, 1998a). We later broadened our view and, taking into account factorial models of differential psychology, extended it to the topic of face and personality impressions (Henss, 1998a). In our empirical studies, we have almost always taken attractiveness into account and in principal component analyses we have always obtained an orthogonal attractiveness factor. This means that attractiveness constitutes an independent personality dimension. The reason we mention this here is the following.

In the 1970s and 1980s, the early days of attractiveness research, the attractiveness stereotype „What is beautiful is good“ was the dominant theme. For some time, the idea prevailed that attractiveness permeates the entire personality assessment, so that basically one evaluative aspect – physical attractiveness – is of overpowering importance and the assessment of people is otherwise rather undifferentiated. Interestingly, Dion, Berscheid and Walster (1972) combined 14 items into a „Social Desirability Index“, i.e. a purely evaluative measure, which was the focus of attention for several years. However, this form of black-and-white painting has been overcome by a more differentiated view (Dermer and Thiel, 1975; Eagly, Ashmore, Makhijani and Longo, 1991; Feingold, 1992). There is no doubt that attractiveness plays a prominent role in face perception and that it also has tangible consequences in various areas of real life, making it one of the most important psychological variables after intelligence. However, there is also no doubt that the human personality is much more multifaceted and that this multiplicity can be reflected in a nuanced assessment of other people's faces – one just has to ask the appropriate questions.

External variables and personality impressions

For the construction of the personality impressions space, we only considered the psychological traits. To examine the relationship with the external variables, we calculated simple correlations, partial correlations controlling for Age, and multiple regressions.

In the multiple regression, we considered each external variable as a criterion and the principal components of the personality space as predictors. Of course, this is not to say that the psychological traits are primary, from which the external variables are inferred. On the contrary, the photographs provide more or less clear or even unambiguous cues for the external variables, while the psychological characteristics can only be assessed

subjectively. The only exception is body height, for which there are hardly any clues in the headshots. We are not concerned with cause and effect, but with correlative relationships, and here we want to focus on the external variables. But first, it should be reminded that the relations shift a bit if one considers the personality factors as the criterion and the external variables as predictors.

Among the external variables, we find a clear ranking. At the top are Age and Smile. Seventy percent of the variance (adjusted) can be explained by the five-dimensional personality impressions space. In second place, but clearly behind, are Glasses and Height. Here, the variance explanation is 40 and 34 percent respectively. For these four variables, the multiple correlation is significant at the 0.1 percent level and in addition, the relationship can be replicated in a cross-validation. For Tilt absolute, Beard, BMI, Hair Loss and Weight, the p-values range from.004 to.040 and the variance explanation from 20 to 12 percent, but the relationship cannot be replicated with the split-half method.

The overriding importance of Age is fully in line with expectations. In the segment considered here, from the late 20s to the late 50s, the age-related changes to the face are considerable and the age estimates show an extraordinarily high level of agreement, even in single stimulus assessments. The very high variance explanation of 70 percent is mainly due to Attractiveness(-) and Masculinity. Although the simple correlations with Agreeableness and with Status fall short of the 5 percent significance level, both generate a small but significant increase in the multiple regression.

The prominent role of Smile is also no surprise. The degree of smiling is clearly visible in photos and it has a close relationship with Negative Affects and the Agreeableness component, on which cheerful, in a good mood and merry have their main loading. Attractiveness also makes a notable contribution to the variance explanation. The increase due to the inclusion of Masculinity is statistically significant, but negligible. The negative correlation with Attractiveness is remarkable. This is apparently an artifact of our sample. The few men who score above the scale mean in the attractiveness ratings and far exceed the others show no trace of smiling. As Figure 6 in the supplement shows, the correlation disappears when these are removed. A comment on Agreeableness is also appropriate. Here there is a positive correlation only up to level 2 on the scale from 0 to 3, after which there is a slight drop (Figure 6 in the supplement). It remains to be seen whether this is an artifact of our sample. In contrast, the negative correlation with Negative Affects is clearly linear, i.e. smiling has a consistently desirable effect here.

Wearing glasses, unsurprisingly, gives the impression of higher professional status and greater intelligence and education. On the other hand, men with glasses appear less masculine. Attractiveness also makes a significant negative contribution in the multiple regression, but the simple correlation is not significant (-.21). The Glasses variable raises two problems. Firstly, it is a dichotomous variable and we did not differentiate it

further.[12] Secondly, and more importantly, only 17 of the 59 men wear glasses.[13] Thus, the risk of selection bias is high. However, the close positive correlation with Status is highly plausible. It is not only the case that people associate glasses with reading, but there is also a genuine link between intelligence and myopia (Davies et al., 2018; Verma and Verma, 2015) and therefore also between intelligence and wearing glasses.

For Height, just over a third of the variance is explained by the personality impressions space. Agreeableness accounts for the largest share and in this respect, shorter men perform better than taller men. When it comes to Attractiveness and Masculinity, however, the advantage lies with the taller men, as to be expected. For these two factors, we observe an interesting connection with Age. The simple correlation between Height on the one hand and Attractiveness and Masculinity on the other is the same (.33 and.32). After controlling for Age, the correlation with Attractiveness is no longer significant (.20), while the correlation with Masculinity soars to.61. The latter indicates that a taller stature is a strong signal of masculinity, which is partially obscured in a sample with a broad range of ages. In contrast, the relationship between Height and Attractiveness is amplified by the age differences. We have not noticed this phenomenon so far (we have not paid attention to it) and it is an interesting question whether this is just an artifact of our sample or whether it has some general validity.

For the other external variables, the relationship with the impressions space is much weaker and not replicable. The already very weak relationship with Tie and Hair Color disappears after controlling for Age. Weight correlates positively with Masculinity, the Body Mass Index negatively with Attractiveness. Both relationships are in line with expectations. The negative correlation between Beard and Status is immediately intelligible, one only has to reflect on the very low proportion of beard wearers in leading positions in business, politics, media and the sciences. A somewhat larger proportion of bearded men with high status are most likely to be found in show business and they usually only wear a stubble beard and often only temporarily. High-status men with a full beard will take a long time to find these days. In our sample, only eight men wear a beard (and thirteen a moustache), so the conclusiveness is limited from the outset. This is all the more true as there are quite different beard styles and these can have different effects depending on the individual face type, which is certainly one of the reasons for the rather inconsistent findings on the impact of facial hair (Dixson and Brooks, 2013; Dixson and Vasey, 2012; Povoa et al., 2024). The absolute head Tilt shows a positive correlation with Status, which is maintained even after controlling for Age (.38,.36). We would not have expected this and it remains to be seen whether this is a peculiarity of our sample. That head tilt can play a role at all, on the other hand, is not unexpected. For example, Vernon et al. (2014, p. 6) report a correlation of.19 with Approachability and.20 with Youthful-Attractiveness. The two factors probably closely correspond

to our Agreeableness and Attractiveness factors, but in our sample there is no correlation with head tilt; the simple correlation is.05 and -.05 respectively, and.01 and.07 after controlling for Age.

This leaves Hair Loss, which was our initial starting point. At the level of the 1618 participants, Hair Loss has a resounding effect, but at the level of the stimulus persons there is only a simple correlation with Attractiveness of -.38, which is no longer significant after taking Age into account (-.22). For those affected, the level of the stimulus persons is certainly the more relevant. This is the place for an important remark. The fact that the amount of cranial hair does not correlate with other personality areas does not mean that it is of no importance. However, in order to demonstrate this, an experimental approach is required in which the hair status is systematically manipulated. In two studies from our laboratory (Becker, 2003; Henss, 2024b), the stimulus persons were naturally bald men who owned a high-quality toupee individually custom-made for them. In experimental online studies in German and English versions, the men were presented either bald or with their toupee. In Henss (2024b), they were rated by independent samples according to attractiveness, self-assurance or health, in Becker (2003) according to intelligence, good husband and family man, successful in his job, or aggressiveness. In terms of attractiveness, eleven out of thirteen men scored significantly better when they wore their toupee and one was significantly more attractive when bald. In terms of intelligence, seven out of fifteen men scored significantly better when seen bald and none with a toupee. Of special interest are the other variables. Here there were some men who scored significantly better with a full head of hair and some who scored significantly better with a bald head, and some for whom there was no difference. This means: Hair status had substantial implications, but because the difference points in this direction for some individuals and in the opposite direction for others, the effects cancel each other out, giving the misleading impression that the abundance of hair is of no importance. The opposing trends make it clear that it is the individual that ultimately plays a decisive role; and this is also then important when there is no systematic relationship in the overall sample.

Typological perspective

Our considerations so far have been entirely in line with conventional research on face perception in social psychology. Now we come to our second perspective, which surprisingly plays almost no role in this field.

Besides the question of consensus between the judges, the focus has so far been on the structural relationships between the variables. The stimulus persons have merely been the carriers of these characteristics and the individuals have only been visible as dots in the Agreeableness / Attractiveness plane in Figure 1. Now we bring the stimulus persons into focus. However, we are not looking at single individuals, but at groups of persons who are similar to each other with regard to a number of characteristics and thus differ from other groups. This is the classical perspective of the typological approach.

To identify person types, we took into account the personality dimensions Agreeableness, Attractiveness, Masculinity, Status and Negative Affects as well as the external variables Smile, Age, Height, Weight, BMI, Hair Loss, Hair Color, Beard, Glasses, Tie and Tilt absolute and subjected these data to a hierarchical cluster analysis. Just as there is no single true factor structure for dimensional analyses, there is also no definite solution for cluster analyses. In order to obtain a fine-grained differentiation, we opted for a 10-cluster solution, which yielded concise differences between the clusters. As an example, we will only look at the four clusters that mark the extremes on the attractiveness dimension and whose personality profiles are shown in Figure 5.

Firstly, we have a group of four men who are far more attractive than the others. This group is the youngest (30 – 37 years) and shows the least degree of hair loss. The men have the darkest hair and they wear neither facial hair nor glasses. The men are estimated to be the second tallest and also rank second in status. In terms of absolute head tilt, they are in second last place, i.e. they hold their heads almost vertically. None of these men displays a smile.

The second most attractive group also contains no spectacle wearers and is the second youngest (30 – 44), but otherwise has a quite different profile. The men have the second lowest body weight and the lowest body mass index. They have the second-lightest hair, show the third highest degree of hair loss and none of them wears a tie. They are in third-last place for status, second-last for masculinity and in terms of social agreeableness, they bring up the rear by a wide margin.

The least attractive group (age 46 – 52) is both the heaviest and the shortest and thus has the highest body mass index. However, it must be emphasized that the men are by no means fat. The weight estimates range from 79.7 to 87.9 kilograms and the BMI of 25.5 to 29.0 lies within the normal range for men in this age group. These men have the second highest degree of hair loss. They display the broadest smiles and score best in terms of mood.

The second least attractive group is the opposite of the least attractive in terms of corpulence. It has the lowest weight (71.0 – 77.2) and the second lowest body mass index (23.3 – 23.8). It scores second best in terms of mood, is the second oldest (44 – 56) and has the highest proportion of tie wearers and no beard wearers.

Even these four brief sketches, in which we have only mentioned the most salient characteristics, make it clear that these are very different types that would not be confused with one another. This of course applies to the contrast between the two most attractive groups on the one hand and the two least attractive groups on the other. But not only that, it also applies to the difference between the two least attractive groups and, to an even greater extent, to the difference between the two most attractive groups.

The contrast between the two most attractive groups and the two least attractive groups gives the following picture. The attractive groups are on average 2.20 points more attractive. Since we are looking at factor scores, this corresponds to 2.2 standard deviation units. The attractive groups are 14.6 years younger, they are estimated to be 3.2 centimeters taller and none of the men wears glasses. Among the unattractive, half wear glasses and the degree of hair loss is 1.38 points stronger on the 7-point Norwood scale. Of particular interest is the difference in smiles. While the attractive score only 0.1 points on the scale from 0 to 3, the unattractive score 1.92. This is certainly one of the main reasons why the unattractive score 0.94 standard deviations better than the attractive in terms of social agreeableness and 0.83 standard deviations better in terms of negative affects. With regard to our other characteristics, there is no notable difference. Roughly speaking, on the one hand we have men who are much more sexually attractive than the others, but at the same time appear less socially agreeable and unsuitable for the family sphere and score much lower in terms of mood. And on the other hand, we have men for whom the opposite is true. The decisive external characteristics are age and the associated features of hair loss and glasses, as well as height and smile. The importance of the smile should be emphasized in particular. It is the most powerful characteristic in cluster formation and, via the items cheerful, in a good mood, merry and sad, it is reflected in both the dimensions of Agreeableness and Negative Affects. Remarkably, of the eleven men in the three most attractive clusters, eight show no trace of smiling, two score 0.2 on the 3-point scale and one scores 0.4. However, there are some attractive men in the other clusters who show a broad smile, and it should be remembered that the negative correlation (-.27) between Smile and Attractiveness is no longer significant after controlling for Age. So one cannot say that sexually attractive men do not smile, but our data are at least an indication that smiling does not increase men's sexual attractiveness. The latter is consistent, for example, with the finding that „a large gender difference emerged in the sexual attractiveness of happy displays: happiness was the most attractive female emotion expression, and one of the least attractive in males" (Tracy and Beall, 2011, p. 1379). And overall, our analysis is also in line with the finding that „smiling enhanced the male facial attractiveness for long-term relationships but not for short-term relationships" (Okubo, Ishikawa, Kobayashi, Laeng and Tommasi, 2015, p. 4).

The fact that the most attractive and least attractive clusters differ substantially from one another is trivial, much more interesting are the differences between neighboring clusters. As mentioned, the two least attractive groups differ fundamentally in terms of corpulence. The least attractive group is the heaviest, the second least attractive the lightest. In addition, there is another remarkable difference. Although the two groups score best on Negative Affects, the difference is 1.24 standard deviation units, as the least attractive group scores best by a very large margin (-1.45 vs. -.22). This is paralleled in smiling: The two groups are ranked first and third, but the difference is 0.88 points (2.46 vs. 1.48; scale 0 – 3).

There are also major differences between the two most attractive groups. Remarkably, the biggest difference can be found for the Attractiveness factor. With a factor score of 2.44, the top group is 1.96 standard deviations above the second most attractive group. While the top group performs best in terms of hair fullness and has the darkest hair, the second most attractive group shows the third most severe hair loss and the second lightest hair. In terms of social agreeableness, the second most attractive group is by far the worst, while the most attractive group is not far below the mean (-1.60 vs. -0.27). In terms of status, the most attractive group ranks second, the second most attractive third last (0.71 vs. -0.59).

These comparisons highlight a very important fact. Attractiveness is a one-dimensional construct and the clusters can be placed in a clear sequence on this axis, but nevertheless neither the attractive faces are all the same, nor the unattractive ones, nor the average attractive ones. On each segment of the attractiveness continuum, there are different types that differ from each other in a variety of ways in terms of other characteristics. In other words, there is not just a single type of attractive faces, but a number of distinct variants; and the same is true for all segments of the attractiveness scale. In a purely dimensional analysis, this important fact lies outside the field of vision; in the typological analysis, it is the focus of attention.

Methodological considerations

It should go without saying that we do not prioritize the typological perspective over the dimensional analysis; after all, the dimensions of the personality impressions space are essential raw data for our cluster analysis.[14] We regard the typological perspective primarily as a valuable addition. If, as in figures 8, 9 and 10, clusters are visualized in the personality space, it is given an additional structure and one may recognize connections that would otherwise remain hidden.

When discussing the Valence-Dominance model in particular, we explicitly pointed out the trivial fact that the result of a dimensional analysis depends crucially on which variables are taken into account. Thereby we emphasized that a purely evaluative factor only comes about if the different domains of personality are not represented by a sufficient number of items. In the following, we would like to focus on some problems of the typological approach.

The basis of our hierarchical cluster analysis are standardized variables and these are given equal weighting. This means, for example, that the characteristic of a tie, which only takes up a minute part of the photos, matters just as much as the powerful characteristics of age and smile and the factors of the personality space. With a small number of characteristics, in our case sixteen, even an inconspicuous characteristic can have a strong impact when given the same weighting. For example, removing Tie leaves only two clusters entirely intact, while the others change to a greater or lesser extent. Interestingly, the stable clusters are the most

doi.org/10.32388/6IDC9O

attractive on the one hand and the least attractive on the other. The most attractive cluster remains stable if only nine or eight clusters are formed, the least attractive even down to five. Remarkably, the least attractive cluster remains intact even if Attractiveness is excluded and at least six clusters are formed. This demonstrates that this cluster is in fact not determined by unattractiveness.[15]

Another interesting finding is at the top end of the attractiveness dimension. The four men in the most attractive cluster occupy places 1, 2, 3 and 4. In this case, attractiveness clearly plays a decisive role. In the second most attractive cluster, the rankings are 5, 6, 22 and – it is hard to believe – 58. This means that the second most attractive cluster contains the second least attractive man.[16] His membership to this cluster is due to similarities in numerous other variables.

These examples contain an important lesson: Clusters are not defined by a single variable, but by a bundle of variables. This can lead to such seemingly curious results that the least attractive cluster is stable even if attractiveness is not taken into account at all, regardless of whether six, seven, eight, nine or ten clusters are formed, and that the second most attractive cluster contains the second least attractive man. One should always remember that the members of a cluster are similar to each other in several respects, but in some other respects they may be quite different.

These findings draw attention to the sample of faces. In this respect, there is a fundamental difference between the dimensional and the typological approach.

The outcome of factor analyses is almost entirely determined by the set of characteristics. The Oosterhof-Todorov strategy enforces an overpowering valence factor, whereas a systematic consideration of different personality domains through a sufficient number of items will largely yield a structure that is known from self-assessments or acquaintance ratings. The face sample only plays a role when narrowly delimited populations are considered. For example, it will be difficult to find a youthfulness factor if only faces of adolescents or very old people are presented.

In typological analysis, on the other hand, the face sample is of utmost importance. A type can only be found if it is represented by several faces. If, like us, one considers a fine-grained resolution, the addition or removal of a single individual may alter the composition of several clusters. A search for generalizable types is therefore many times more difficult than a search for personality dimensions. The latter have essentially been known for a long time. All one has to do here is make use of the treasure trove of differential psychology and one is only faced with the problem of which domains to consider. When it comes to the question of face types, however, we are at the very beginning.

Brief summary

To conclude, we would first like to recall our starting point. Our original research questions focused on the social perception of genetic hair loss in men. To this end, we took into account an age range in which the genetic predisposition manifests itself more and more strongly, and we compiled the personality-descriptive items against the background of our previous studies. Our aim was neither to identify face types nor to take comprehensive account of the multifaceted personality impressions. The results of the current study are predetermined by these constraints.

In the dimensional analysis, we obtained exactly what is to be expected based on the state of research and what we have repeatedly confirmed in numerous studies of our own. The perception of unfamiliar faces is not a black-and-white painting, but remarkably differentiated and the structure of personality impressions is essentially a reflection of the personality structure that has long been known in differential psychology from self-assessments and assessments of acquaintances. This is actually an extraordinary fact, but one that is rarely emphasized explicitly and of which some researchers are probably not aware. Due to the close correspondence between assessments of strangers, self and acquaintances, it is not to be expected that studies on face perception will reveal completely new personality factors.

With the typological perspective, however, we have entered new territory. To be more precise, new territory in relation to men, because in the 1990s we conducted several studies on types of attractive women.[17] The interest in a typological view on men's faces came mainly from the observation that, for different characteristics, baldness is a disadvantage for some men but an advantage for others, and that hair fullness makes no significant difference for some men. On the one hand, this highlights the importance of the individual physiognomic configuration, but it also suggests that similar faces can be grouped into types that evoke similar personality impressions in several respects.

In our analysis, we did not start from the superordinate concept of similarity, but used the available assessments and also took into account some external variables that we considered potentially relevant. By linking the psychological personality space with the cluster structure, we gained insights that would have remained hidden in a purely dimensional view. We consider the linking of the two perspectives to be the most valuable contribution of this study. The extent to which the different types can be generalized is an open question. Here we have merely given an impulse that requires further research.

Conclusions

The human face is the most meaningful object in our social world. People are effortlessly able to draw conclusions about the most diverse characteristics of the person in question simply by looking at another

person's face. Although the judgments are subjective, there is a certain inter-individual consensus that varies depending on the characteristic. The impressions are not unconnected and unsystematic, but have a clear structure that can be represented as a spatial model in a dimensional analysis using factor analytic methods. An all-powerful evaluation factor, as in the popular Valence-Dominance Model, only emerges if different domains are considered with too few or even just a single item. If, on the other hand, different domains are each represented with a sufficient number of items, a highly differentiated structure is obtained, which is by and large the same as has long been known in differential psychology on the basis of self- and acquaintance evaluations. From a typological perspective, cluster analytic methods can be used to identify groups of people who are similar to each other in a bundle of characteristics and are clearly different from other groups. The typological approach is extremely rare in social psychological face research, but on the one hand, it provides unique insights on its own and on the other hand, combining the dimensional and typological perspectives facilitates insights that are more than „the sum" of the individual approaches.

Footnotes

[1] Jones et al. (2021) tested the appropriateness of the Valence-Dominance model with 11,148 participants from 41 countries from the regions of Africa, Asia, Australia and New Zealand, Central America and Mexico, Eastern Europe, the Middle East, the USA and Canada, Scandinavia, South America, the UK, and Western Europe. The stimulus material consisted of photographs of 15 male and 15 female faces each of Whites, Blacks, Asians and Latinos. Their „results suggest the valence-dominance model generalizes across world regions when using an identical analysis to Oosterhof and Todorov's original study" (p. 23). Additional exploratory factor analyses (EFA), on the other hand, showed „little evidence that the valence-dominance model generalizes across world regions" (p. 24). In their comparison with the Jones et al. data, Todorov and Oh (2021, pp. 204 and 207) come to the conclusion „that the structure of judgments derived from both a PCA and an EFA is remarkably consistent across cultures... the first PC derived from the PCA... would be considered structurally identical in all world regions... the second PC... would be considered identical in nine world regions and very similar in the remaining two regions (Asia and the Middle East)... The structure of judgments derived from the EFA is also remarkably consistent across cultures".

[2] The item dominant is the only one to correlate substantially with only two others (temperamental.72 and masculine.73). Since masculine correlates much more strongly with attractive (-.82) and also with physically healthy (-.69) and extravert (-.62), and the three in turn correlate closely with each other, masculine belongs to this group and not to dominant.

[3] A comprehensive description can be found in our book „Gesicht und Persönlichkeitseindruck [Face and Personality Impressions]“, Henss (1998a).

[4] Below, we will present a five-factor solution at the level of the stimulus persons. The comparison with the five-factor solution at the level of the judges using Tucker’s congruence analysis yields the ϕ-coefficients.98,.96,.94,.76 and.73. This means that three components are very similar, whereas two components only match coarsely. As ϕ corresponds to the cosine of the angle between the components, in the worst case the angle is 43.1 degrees. However, if one interprets the components only according to the main loadings, one could give the respective ones the same label. In other words, the core is essentially the same despite the poor ϕ values.

[5] On the close relationship between Extraversion and Positive Affects, see for example Watson and Clark (1997).

[6] This is not a congruence analysis, in which the agreement of the loadings is determined. This would not make sense when comparing an unrotated and a rotated solution.

[7] Originally, the loadings have the opposite sign. We have reversed them according to their desirability in men.

[8] Although this component has similarities, it is not the same as Negative Affects in the Positive and Negative Affect Scale PANAS by Watson, Clark and Tellegen (1988).

[9] In the cross-cultural study by Hester, Xi and Hehman (2021), 11,481 participants from 45 countries assessed 120 faces on the 13 adjectives of Oosterhof and Todorov. Surprisingly, emotionally stable was ranked 6th in terms of consensus among judges. However, the gap to last-placed intelligent is not very large (ICC.095 vs..077). At the top is attractive with.151. The same can also be found in the supplement to Oosterhof and Todorov (2008).

[10] In our case, too, cheerful, merry and in a good mood on the one hand and nervous, insecure and anxious on the other hand are almost perfectly orthogonal.

[11] The angle between the item dominant and the principal component Agreeableness is almost exactly 45 degrees.

[12] Glasses are not all the same. For example, Leder, Forster and Gerger (2011) have shown that full-rim and rimless glasses can have different effects.

[13] According to a survey conducted by the Allensbach Institute for the Zentralverband der Augenoptiker und Optometristen [Central Association of Opticians and Optometrists], 63 percent of the German male population

aged 16 and older wore glasses in 2019; the figure for women was even higher at 70 percent (ZVA, 2019). This means that spectacle wearers are heavily underrepresented in our sample.

[14] A dimensional analysis is not a necessary prerequisite for a cluster analysis. The input data for a cluster analysis can also be obtained via similarity sorting, for example. Here, the participants are asked to sort a set of faces into groups according to similarity, whereby it is up to them to decide what they understand by similarity and how many clusters they form.

[15] The five men are ranked 28th, 39th, 40th, 56th and 59th.

[16] This is a very strange case. We would have placed this man in the middle, definitely not in the bottom quartile.

[17] With different face samples, we used similarity sorting to determine types of attractive women and found the following, for example (Bäsel, 2003, Henss, 1997b, 1998a, c, 1999). Participants have no problems assembling different types of women. There is a high level of consensus in the assignment of labels such as 'Classical Beauty', 'Vamp', 'Girl Next Door', 'Lolitas' to the corresponding groups. The types have quite different personality profiles. A high proportion of the participants show consistent preferences for certain types. There are some, albeit rather weak, correlations between self-assessed personality traits and type preferences. For other studies on types of attractive women, see for example Ashmore, Solomon und Longo (1996), Berry (1991), Solomon, Ashmore und Longo (1992); for general considerations on types of attractive men and women see Marwick (1988).

References

- Ashmore, R.D., Solomon, M.R. & and Longo, L.C. (1992). Thinking about fashion models' looks: A multidimensional approach to the structure of perceived physical attractiveness. *Personality and Social Psychology Bulletin*, 22, 1083-1104.
- Bäsel, H.H. (2003). *Schöne Frauen sind nicht alle gleich. Urteilerpersönlichkeit und Typenpräferenzen bei der Beurteilung attraktiver Frauengesichter*. Unpublished Diploma Thesis at the Fachrichtung Psychologie, University of the Saarland, Saarbrücken.
- Bar, M., Neta, M. & Linz, H. (2006). Very first impressions. *Emotion,* 6 (2), 269-278. https://doi.org/10.1037/1528-3542.6.2.269
- Becker, S. (2003). *Zur Eindruckswirkung des Haarausfalls beim Mann*. Unpublished Diploma Thesis at the Fachrichtung Psychologie, University of the Saarland, Saarbrücken.

- Berry, D. S. (1991). Attractive faces are not all created equal: Joint effects of facial babyishness and attractiveness on social perception. *Personality and Social Psychology Bulletin*, 17 (5), 523-531. https://doi.org/10.1177/0146167291175007
- Blaker, N.M., Rompa, I., Dessing, I.H., Vriend, A.F., Herschberg, C. & van Vugt, M. (2013). The height leadership advantage in men and women: Testing evolutionary psychology predictions about the perceptions of tall leaders. *Group Processes & Intergroup Relations*, 16, 17-27. https://doi.org/10.1177/1368430212437211
- Borkenau, P., Brecke, S., Möttig, C. & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, 43 (4), 703-706. doi:10.1016/j.jrp.2009.03.007
- Borkenau, P. & Liebler, A. (1992). The cross-modal consistency of personality: Inferring strangers' traits from visual or acoustic information. *Journal of Research in Personality*, 26, 183-204.
- Case, A. & Paxson, C. (2008). Stature and status: Height, ability, and labor market outcomes. *Journal of Political Economy*, 116 (3), 499-532. https://doi.org/10.1086/589524
- Chan, H.F., Skali, A., Stadelmann, D., Torgler, B. & Whyte, S. (2020). Masculinity cues, perceptions of politician attributes, and political behavior. *Economics & Politics*, 2020; 00, 1-24. https://doi.org/10.1111/ecpo.12164
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. San Diego: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Davies, G., Lam, M., Harris, S.E. et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications*, 9, 2098. https://doi.org/10.1038/s41467-018-04362-x
- Dermer, M. & Thiel, D.L. (1975). When beauty may fail. *Journal of Personality and Social Psychology*, 31, 1168-1176.
- Dion, K., Berscheid, E. & Walster, E. (1972). What is beautiful is good? *Journal of Personality and Social Psychology*, 24, 285-290.
- Dixson, B. J. & Brooks, R.C. (2013). The role of facial hair in women's perceptions of men's attractiveness, health, masculinity and parenting abilities. *Evolution and Human Behavior*, 34 (3), 236-241. https://doi.org/10.1016/j.evolhumbehav.2013.02.003
- Dixson, B.J & Vasey, P.L. (2012). Beards augment perceptions of men's age, social status, and aggressiveness, but not attractiveness, *Behavioral Ecology*, 23 (3), 481-490. https://doi.org/10.1093/beheco/arr214
- Eagly, A.H., Ashmore, R.D., Makhijani, M.G. & Longo, L.C. (1991). What is beautiful is good, but: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110, 109-128.
- Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, 111, 304-341.

- Goldstein, A.G. & Papageorge, J. (1980). Judgments of facial attractiveness in the absence of eye movements. *Bulletin of the Psychonomic Society*, 15, 269-270.
- Guthrie, R. D. (1976). *Body hot spots. The anatomy of human social organs and behavior*. New York: Van Nostrand Reinhold Company.
- Harper, B. (2000). Beauty, stature and the labour market: A British cohort study. *Oxford Bulletin of Economics and Statistics*, 62, 771-800. https://doi.org/10.1111/1468-0084.0620s1771
- Harris, M.A., Brett, C.E., Deary, I.J. & Starr, J.M. (2016). Associations among height, body mass index and intelligence from age 11 to age 78 years. *BMC Geriatrics*, 16:167. https://doi.org/10.1186/s12877-016-0340-0
- Hehman, E., Sutherland, C.A.M., Flake, J.K. & Slepian, M.L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, 113 (4), 513-529. https://doi.org/10.1037/pspa0000090
- Henss, R. (1987). Zur Beurteilerübereinstimmung bei der Einschätzung der physischen Attraktivität junger und alter Menschen. *Zeitschrift für Sozialpsychologie*, 18, 118-130.
- Henss, R. (1992). *„Spieglein, Spieglein an der Wand...“ Geschlecht, Alter und physische Attraktivität*. Weinheim: Psychologie Verlags Union.
- Henss, R. (1995). From Aal to Zyniker. Personality-descriptive type nouns in the German language. *European Journal of Personality*, 9, 125-135.
- Henss, R. (1996). Fantasy, wishful thinking, and statistical hokuspokus. Rejoinder to De Raad and Ostendorf (1996). *European Journal of Personality*, 10, 157-160.
- Henss, R. (1997a). Face and personality impressions. Consensus at zero acquaintance. Poster presented at the *4. Arbeitstagung Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik*. University of Bamberg, Germany. Oct. 30 - Nov. 01, 1997. https://doi.org/10.13140/RG.2.2.11895.23205
- Henss, R. (1997b). *Schönheit hat viele Gesichter. Typen attraktiver Frauen*. Presentation at the 6. Tagung der Fachgruppe Sozialpsychologie. 20.-22. Juni, 1997. Konstanz.
- Henss, R. (1998a). *Gesicht und Persönlichkeitseindruck*. Göttingen: Hogrefe. [Reprint: Henss, R. (2020). *Gesicht und Persönlichkeitseindruck*. Saarbrücken: Dr. Ronald Henss Verlag. Freely available at https://www.researchgate.net/publication/339147401_Gesicht_und_Personlichkeitseindruck
- Henss, R. (1998b). Type nouns and the five factor model of personality description. *European Journal of Personality*, 12, 57-71.
- Henss, R. (1998c). *Schöne Menschen sind nicht alle gleich. Zur Differenzierung des Attraktivitätsstereotyps*. Poster presented at the 41. Kongress der Deutschen Gesellschaft für Psychologie. Dresden, 27.09. - 01.10.1998.
- Henss, R. (1999). *Attraktive Gesichter. Typenpräferenz und Urteilerpersönlichkeit*. Presentation at the 7. Tagung der Fachgruppe Sozialpsychologie. Kassel. 10. - 13. Juni 1999.

- Henss, R. (2001). Social perceptions of male pattern baldness. A review. *Dermatology and Psychosomatics / Dermatologie und Psychosomatik*, 2 (2), 63-71. https://www.researchgate.net/publication/244915177_Social_Perceptions_of_Male_Pattern_Baldness_A_Review
- Henss, R. (2017). *Die Körpergröße deutscher Männer von 1896 bis 1996*. https://splitter1.wordpress.com/2017/07/15/koerpergroesse-maenner-deutschland/
- Henss, R. (2024a). Hair loss, body height and attractiveness malus for men. *Qeios*. https://doi.org/10.32388/NIAXZ7
- Henss, R. (2024b). „Toupee or not Toupee?" Sic (et Non). Social perceptions of Male Pattern Baldness (Androgenetic Alopecia). *Qeios*, https://doi.org/10.32388/UYKBF4
- Henss, R. (2024c). Blue Screen – Blue Mood? Influence of background color and attractiveness of female stimulus persons on current mood in an online experiment (PANAS). Qeios, https://www.qeios.com/read/FEVW0Z.2
- Hester, N., Xie, S.Y. & Hehman, E. (2021). Little between-region and between-country variance when people form impressions of others. *Psychological Science*, 32 (12), 1907-1917. https://doi.org/10.1177/09567976211019950
- Jones, B.C., DeBruine, L.M., Flake, J.K. et al. (2021). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behavior*, 5, 159-169. https://doi.org/10.1038/s41562-020-01007-2
- Judge, T.A. & Cable, D.M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. Journal of Applied Psychology, 89 (3), 428-41. https://doi.org/10.1037/0021-9010.89.3.428
- Koo, T.K. & Li, M.Y. (2016). A guideline of selecting and reporting Intraclass Correlation Coefficients for reliability research. *Journal of Chiropractic Medicine*, 5 (2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012
- Kurz, K. & Register, J. (1998). *Auf den allerersten Blick: Urteilerübereinstimmung bei kurzzeitiger Darbietung von Gesichterfotos*. Unpublished Diploma Thesis at the Fachrichtung Psychologie, University of the Saarland, Saarbrücken.
- Leder, H., Forster, M. & Gerger, G. (2011). The glasses stereotype revisited. Effects of eyeglasses on perception, recognition, and impression of faces. *Swiss Journal of Psychology*, 70 (4), 211-222.
- Lin, C., Keleş, U. & Adolphs, R. (2019). Four dimensions characterize attributions from faces using a representative set of English trait words. *Nature Communications*, 12. https://www.nature.com/articles/s41467-021-25500-y.pdf
- Locher, P., Unger, R., Sociedade, P. & Wahl, J. (1993). At first glance: Accessibility of the physical attractiveness stereotype. *Sex Roles*, 28, 729-743.

- Marwick, A. (1988). *Beauty in history. Society, politics and personal appearance c. 1500 to the present*. London: Thames and Hudson.
- Montepare, J.M. & Zebrowitz, L.A. (1998). Person perception comes of age: The salience and significance of age in social judgments. *Advances in Experimental Social Psychology*, 39, 93-161.
- Muscarella, F. & Cunningham, M.R. (1996). The evolutionary significance and social perception of male pattern baldness and facial hair. *Ethology and Sociobiology*, 17, 99-117.
- NCD (2016). *A century of trends in adult human height*. NCD Risk Factor Collaboration (NCD-RisC). https://doi.org/10.7554/eLife.13410
- Norwood, O'T.T. (1975). Male pattern baldness: Classification and incidence. *Southern Medical Journal*, 68, 1359-1365.
- Okubo, M., Ishikawa, K., Kobayashi, A., Laeng, B. & Tommasi, L. (2015). Cool guys and warm husbands: The effect of smiling on male facial attractiveness for short- and long-term relationships. *Evolutionary Psychology*, 13 (3). https://doi.org/10.1177/1474704915600567
- Oosterhof, N.N. & Todorov, A. (2008). The functional basis of face evaluation. *PNAS*, 105 (32), 11087-11092. https://doi.org/10.1073/pnas.0805664105
- Osgood, C.E. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, 12, 194-199.
- Osgood, C.E., Suci, G. & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana-Champaign: University of Illinois Press.
- Pisanski, K., Fernandez-Alonso, M., Díaz-Simón, N., Oleszkiewicz, A., Sardinas, A., Pellegrino, R., Estevez, N., Mora, E.C., Luckett, C.R. & Feinberg, D.R. (2022). Assortative mate preferences for height across short-term and long-term relationship contexts in a cross-cultural sample. *Frontiers in Psychology*, 13:937146. https://doi.org/10.3389/fpsyg.2022.937146
- Povoa, A., Pech, W., Korndorfer, H.B., Viacava, J.C. & Jacinto, L.C. (2024). Is beard the male makeup? An experimental study on trust perception based on appearance. SSRN. http://dx.doi.org/10.2139/ssrn.4838386
- Roberts, J.V. & Herman, C.P. (1986). The psychology of height: An empirical review. In Herman, C.P., Zanna, M.P. & Higgins, E.T. (Eds.). *Physical appearance, stigma, and social behavior. The Ontario symposium* (Vol. 3, 113-140). Hillsdale, NJ. Erlbaum.
- Schmidt, K. (1999). *Auf den allerersten Blick: Persönlichkeitsbeurteilung bei minimaler Beobachtungszeit*. Unpublished Diploma Thesis at the Fachrichtung Psychologie, University of the Saarland, Saarbrücken.
- Shrout, P.E. & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86* (2), 420-428. https://doi.org/10.1037/0033-2909.86.2.420

- Shrout, P.E. & Lane, S.P. (2012). Reliability. In Cooper, H., Camic, P.M., Long, D.L., Panter, A.T., Rindskopf, D. & Sher, K.J. (Eds.), *APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics*, 643–660. American Psychological Association. https://doi.org/10.1037/13619-034
- Sutherland, C.A.M., Oldmeadow, J.A., Santos, I.M., Towler, J., Burt, D.M. & Young, A.W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, 127, 105–118. https://doi.org/10.1016/j.cognition.2012.12.001
- Todorov, A. (2017). *Face value. The irresistible influence of first impressions*. Princeton, NJ: Princeton University Press.
- Tracy, J.L. & Beall, A.T. (2011). Happy guys finish last: The impact of emotion expressions on sexual attraction. *Emotion,* 11 (6), 1379-1387. https://doi.org/10.1037/a0022902
- Tyrrell, J., Jones, S.E., Beaumont, R., Astley, C.M., Lovell, R., Yaghootkar, H. et al. (2016). Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *BMJ*, 352:i582. https://doi.org/10.1136/bmj.i582
- Verma, A. & Verma, A. (2015). A novel review of the evidence linking myopia and high intelligence. *Journal of Ophthalmology*, Article ID 271746. http://dx.doi.org/10.1155/2015/271746
- Vernon, R.J.W., Sutherland, C.A.M., Young, A.W. & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *PNAS*, 111 (32), E3353-E3361. https://doi.org/10.1073/pnas.1409860111
- Vuoksimaa, E., Panizzon, M.S., Franz, C.E. et al. (2018). Brain structure mediates the association between height and cognitive ability. *Brain Structure and Function*, 223, 3487-3494. https://doi.org/10.1007/s00429-018-1675-4
- Walker, M. & Vetter, T. (2015). Changing the personality of a face: Perceived Big Two and Big Five personality factors modeled in real photographs. *Journal of Personality and Social Psychology,* 110 (4), 609-624. https://doi.org/10.1037/pspp0000064
- Wang, J., Chen, Q., Chen, G., Li, Y., Kong, G. & Zhu, C. (2020). What is creating the height premium? New evidence from a Mendelian randomization analysis in China. *PLoS ONE*, 15 (4): e0230555.
- Watson, D. & Clark, L.A. (1997). Extraversion and its positive emotional core. In Hogan, R., Johnson, J. & Briggs, S. (1997). *Handbook of personality psychology*. San Diego: Academic Press.
- Watson, D., Clark, L.A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Wiggins, J.S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37, 395-412.
- Wiggins, J.S., Phillips, N. & Trapnell, P. (1989). Circular reasoning about interpersonal behavior: Evidence concerning some untested assumptions underlying diagnostic classification. *Journal of Personality and*

Social Psychology, 56, 296-305.

- Wiggins, J.S. & Pincus, A.L. (1992). Personality: Structure and assessment. *Annual Review of Psychology*, 43, 473-504.
- Willis, J. & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17 (7), 592-598. https://doi.org/10.1111/j.1467-9280.2006.01750.x
- Wolffhechel, K, Fagertun, J, Jacobsen, U.P., Majewski, W, Hemmingsen, A.S. et al. (2014) Interpretation of appearance: The effect of facial features on first impressions and personality. *PLoS ONE*, 9 (9): e107721. https://doi.org/10.1371/journal.pone.0107721
- Zebrowitz, L.A. (1997). *Reading faces. Windows to the soul?* Boulder, CO: Westview Press.
- Zebrowitz, L.A. (2011). Ecological and social approaches to face perception. In Calder, A., Haxby, J.V., Johnson, M. & Rhodes, G. (Eds.). *Handbook of face perception*. Oxford: Oxford University Press.
- ZVA (2019). *Brillenstudie 2019*. Düsseldorf: Zentralverband der Optiker und Optometristen. https://www.zva.de/brillenstudie.

Supplementary data: available at https://doi.org/10.32388/6IDC9O

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.