Open Peer Review on Qeios

Blue Screen – Blue Mood? Influence of background color and attractiveness of female stimulus persons on current mood in an online experiment (PANAS)

Ronald Henss

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

In an online experiment, more than 6,000 volunteers rated women's faces according to their likeability The women were either extraordinarily attractive or of medium to low attractiveness. The background of the website was either Black, Blue, Green, Gray, Red or Yellow. Following the likeability assessments, the participants were asked to assess their own current mood using the PANAS (Positive and Negative Affect Schedule). The study was conducted in a German-language and an English-language version. An exploratory factor analysis with Oblimin rotation shows perfect orthogonality for the Positive Affects and Negative Affects (r =.01) and a reliability analysis shows excellent reliability for both scales (Positive Affects: Cronbach's α =.87, McDonald's ω =.88; Negative Affects: α =.88, ω =.89). In the case of Negative Affects, the scores show an extreme concentration at the lower end of the scale. The attractiveness of the stimulus persons has a significant but weak effect on Negative Affects (Cohen's d = 0.06). The background color of the website shows an effect on Positive Affects, but not on Negative Affects. The clear favorite is Green (for Green/Blue d = 0.15). The participants in the English-language version score higher on Negative Affects (d = 0.25). On Positive Affects, males score higher than women (d = 0.07). The older subjects score better on both scales than the younger ones (for age group comparisons, d values up to 0.46). Screen color has no notable effect on the assessment of likeability of the

women's faces.

Keywords: Color, Attractiveness, Female Faces, Mood, Emotional State, PANAS, Positive Affects, Negative Affects, Online-Experiment, Web Design, Human-Computer Interaction.

Highlights

- PANAS Positive Affects and Negative Affects are orthogonal
- PANAS scales highly reliable
- Older subjects higher Positive and lower Negative Affects
- English version somewhat higher on Positive and Negative Affects than German version
- Green website background raises Positive Affects
- No effects of background color on Negative Affects

Introduction

Colors are among the most important factors in our sensory perception and there is no doubt that they can exert an enormous influence in a wide variety of areas of life. In the present study, we look at this topic from a narrowly circumscribed perspective, namely with regard to the question of whether the color design of the background in an online experiment has an influence on the current mood state and the assessment of women's faces. In addition, there is the question of the usefulness of the questionnaire used to measure mood.

In a short online experiment, the participants had to rate photographs of women's faces according to their likeability. The women were either exceptionally attractive or of medium to low attractiveness. The judges were then asked to assess their own current mood. Depending on the test condition, the background of the website was Black, Blue, Green, Gray, Red or Yellow. The study was conducted in a German and an English version.

To assess the current mood state, the Positive and Negative Affect Schedule (PANAS) was used. The short questionnaire developed by Watson, Clark and Tellegen (1988) comprises 20 adjectives that address various aspects of emotional state (see Table 1). Contrary to what the name suggests, the Positive and Negative Affects are not pairs of opposites. Instead, the questionnaire was explicitly designed to measure two independent factors of mood.

Along with the two core variables of color and attractiveness, the language version (German/English) and the sex and age of the participants are also taken into account. As the study is explicitly exploratory in nature, no hypotheses are stated.

Here it is helpful to know the background. The experiment was started at the end of 1999[1]. We began conducting psychological experiments on the Internet in the mid-1990s. Those were the early years of the Internet and there was

criticism from all sides that such experiments would make no sense in the field of psychology. Today it is the most common practice of all. Back then, we taught ourselves and our students how to conduct experiments on the Internet and discovered that they had very different preferences for the color design. This raised the question of whether colors have a significant impact and may need to be systematically controlled as a potentially confounding variable.

Our research focused on the area of face perception, particularly on physical attractiveness (Henss, 1991, 1992, 1993, 1998a), and on differential psychology with an emphasis on the factor-analytic structure of personality and personality impressions (Henss, 1997,1998a, 1998b).

We were also looking for a short instrument to measure mood that we could routinely run alongside our experiments. We chose the PANAS, which had been around for some years, but had not been used as extensively and had not yet been researched as intensively as it is today.[2] This raised the question of how suitable the questionnaire is for German-speaking populations and whether there are differences between the German and English versions.

Against this background, it was obvious to combine the very different aspects of the website's color scheme, attractiveness and mood in an experiment.

In a preliminary version of this article, reviewers raised the question of why the attractiveness of faces was made the subject of research. This question is irritating. People are the most important entities in our world. The human face is by far the most important part of human appearance. Nothing on the human body conveys such diverse and rich information as the face and attractiveness plays a paramount role in the assessment of our fellow human beings.

The proverb „Don't judge a book by its cover" is intended to warn us not to judge people by their outward appearance, but this admonition is as fruitless as it is impracticable. People cannot help but judge their fellow human beings – also, but not only – by their outward appearance. Face judgments have the character of an instinct; they are automatic, unconscious and lightning-fast. Even if a facial photograph is only visible for 40 milliseconds, viewers have impressions of the most diverse aspects of personality, which change only slightly even with longer viewing times. Newborns can distinguish between happy and sad faces after just one day. Infants at the age of a few months have preferences for attractive faces. Our brain contains specialized modules for face perception and even neurons that are attuned to the attractiveness of faces. For these and many more highly interesting facts about face perception, see for example Todorov (2017), Alexander (2024), Langlois et al. (1987, 1991), Landau (1989), Liang, Zebrowitz and Zhang (2010), Rhodes and Zebrowitz (2002), Principe and Langlois (2011), Zebrowitz (1997).

Attractive faces have a very high reward value, not only for infants. Adults also prefer to look at attractive faces rather than average or even unattractive ones. One only has to look at the media, which bombard us day after day and hour after hour with attractive people and pretend a world that doesn't exist.

There is no doubt that we have more positive feelings when we look at attractive faces and that the sight of very unattractive faces arouses negative feelings – whether we like it or not. For us, the question was whether such different emotions would be reflected in a short online experiment on a mood questionnaire.

There is also no doubt that colors arouse different emotions. But here, too, the question arose as to whether this could be demonstrated in an online experiment via a mood questionnaire. It was even more unclear which background color of the screen could produce which effect. Both in terms of mood and in terms of the assessment of likeability of the women's faces.

Overall, the experiment addresses three questions. Firstly, the psychometric quality of the measuring instrument PANAS and its suitability for the German-speaking realm and the comparison with the English version. Secondly, the influence of the attractiveness of the stimulus persons on mood. Thirdly, it is about the effect of the background color of the website, on the one hand on mood and on the other hand on the assessment of the likeability of the women's faces. As mentioned, the investigation had the character of an exploratory experiment and we do not pretend in retrospect that we had specific hypotheses at the time.

Some readers may wonder why we have not mentioned a single source on color psychology. The reason for this will hopefully become clear in the discussion.

Methods

Procedure

Participants were recruited via links on our home page at the psychological institute of the University of the Saarland in Saarbrücken, which was well frequented due to numerous previous online studies. Participation was voluntary, free, non-binding and anonymous. The experiment ran on a single website and only took a couple of minutes.

The key independent variable is the Color that fills the background of the website. The color values Red (RGB 990000), Yellow (FFFF99), Green (009900), Blue (000099), Grey (CCCCCC) or Black (000000) were deliberately chosen to be intense in order to maximize any effects.

Five panels, each with three portrait photos measuring 180 x 240 pixels, were presented one below the other. Within each panel, the participants were asked to mark the face they found most likeable. The stimulus persons were either exceptionally attractive, some of them world-famous supermodels, or of medium to low attractiveness. The pictures were presented in a fixed arrangement. In the second group, attractiveness decreased from panel to panel, so that the last triplet was rather unattractive.[3] Such a gradation was neither possible nor desired for the exceptionally attractive women.

Immediately after the likeability ratings, the judges were asked to assess their own current mood state.

Finally, the participants were asked to state their Sex, Age and Country of Origin or, in the German version, the federal state. In this article, no distinction is made between the countries or regions of origin, but only between the German/English Language version.

This results in the 6 x 2 x 2 x 2 experimental design Color x Attractiveness x Sex x Language; and Age is considered as a

covariate. The core variables Color and Attractiveness are randomized, Sex, Language version and Age inherently result from self-selection.

The PANAS

The current mood was assessed using the PANAS. The scale consists of the 20 items listed in Table 1 in the order used in the experiment (the order is alternating and differs slightly from the customary order).

Table 1. *PANAS Items German and English version.*

German	English	
aktiv	active	PA
bekümmert	distressed	NA
interessiert	interested	PA
verärgert	upset	NA
freudig erregt	excited	PA
schuldig	guilty	NA
stark	strong	PA
erschrocken	scared	NA
angeregt	inspired	PA
feindselig	hostile	NA
stolz	proud	PA
gereizt	irritable	NA
begeistert	enthusiastic	PA
beschämt	ashamed	NA
wach	alert	PA
nervös	nervous	NA
entschlossen	determined	PA
durcheinander	jittery	NA
aufmerksam	attentive	PA
ängstlich	afraid	NA

10 items each measure Positive Affect (PA) and Negative Affect (NA). The assessment was made on a 5-point scale, which was verbally anchored by „not at all“, „a little“, „moderately“, „quite a bit“, „extremely“ („gar nicht“, „ein bisschen“, „einigermaßen“, „erheblich“, „äußerst“). Usually, a score from 1 to 5 is used and the scale value is calculated as the sum, resulting in a range from 10 to 50. We have transformed the value to the interval [0 – 4] using the transformation x/10 - 1. We consider this to be more transparent and, as the lowest scale value was anchored with „not at all“, also more appropriate in terms of the content.

In addition, the two items „in a good mood“ and „cheerful“ were added and the mean value was calculated as a further

measure of positive feelings. We refer to the scale as GM for Good Mood.

As our data base is exceptionally large, all cases with more than one missing value were excluded. In the remaining 6,113 cases, the proportion of missing values is 0.4 percent. There is a peculiar feature in the English version of the „upset“ item. This alone accounts for 291 (52.6%) missing values, compared to only 5 for the German equivalent „verärgert“. The assumption that the respondents were non-native speakers was not confirmed. This phenomenon is interesting, but it cannot have any effect on our findings.

The number of subjects at the levels of the independent variables is summarized in Table 2. Of the 48 cells in the experimental design, only 10 contain fewer than 100 subjects, the minimum being 82. Thus, this study has an unusually high test power.

Table 2. *Number of subjects.*

		N				N
Color	Red	996		Attractiveness	very high	3077
	Black	1021			medium, low	3036
	Green	1007		Sex	Males	2447
	Blue	1043			Females	3666
	Yellow	1000		Language	German	2639
	Grey	1046			English	3474

As is usual with online experiments, the participants were quite young and the majority came from an academic background and the English-language version was dominated by the Anglosphere (USA 64.3 percent; Canada 12.5; UK 6.0, Australia / New Zealand 1.5). The age range extends from 15 to 74 years, the mean is 26 years, the median 23 years, and only 5 percent were older than 46 years. The mean age at the different levels of the independent variables can be seen in Table 3.

Table 3. *Age of subjects.*

		M				M
Color	Red	26,2		Attractiveness	very high	26,0
	Black	25,7			medium, low	25,9
	Green	26,0		Sex	Males	28,6
	Blue	25,8			Females	24,2
	Yellow	26,2		Language	German	26,8
	Grey	25,7			English	25,4

The assignation to color and attractiveness was randomized. As expected, the age differences are statistically

insignificant. For the language version and sex, however, the differences are significant at the 0.1% level. The participants in the German version are 1.43 years older and this corresponds to an effect size (Cohen's d) of 0.15. The men are considerably older than the women; the age difference is 4.42 years, the effect size is 0.48.

Results

Our research questions fall into two different blocks. The first concerns the psychometric properties of the PANAS, the second the effects of the independent variables on the dependent variables.

Factorial structure of the PANAS

First, the suitability of the PANAS was examined. Since the factor structure is clearly predetermined, a confirmatory factor analysis was conducted first. The model with uncorrelated measurement errors shows a poor fit. A model with several correlated errors, on the other hand, results in an excellent fit (CFI and TLI >.95 and SRMR and RMSEA <.05), as can be seen in Table 4.[4]

Table 4. *Confirmatory factor analysis, correlated errors. Goodness of fit.*

				RMSEA 90% CI		Chi square		
CFI	TLI	SRMR	RMSEA	Lower	Upper	χ^2	df	p
0.968	0.955	0.0499	0.0442	0.0424	0.0461	1734	134	<.001

Subsequently, an exploratory factor analysis with Maximum Likelihood extraction and Oblimin rotation was run. This resulted in two eigenvalues greater than 1 (the third largest is 0.65). The first factor (NA) explains 22.8 percent of the variance, the second (PA) 20.8 percent. The KMO criterion for the overall scale is.91. The MSA values of the items range from.86 to.93. KMO and MSA values above.90 are considered excellent. The loadings of the pattern matrix are shown in Table 5. Loadings of <.10 are omitted.

Table 5. *Factor loadings. Pattern matrix, Maximum Likelihood extraction, Oblimin rotation.*

		NA	PA
ängstlich	afraid	0,780	
durcheinander	jittery	0,715	
nervös	nervous	0,711	
erschrocken	scared	0,701	
beschämt	ashamed	0,687	
gereizt	irritable	0,649	
bekümmert	distressed	0,630	
schuldig	guilty	0,629	
feindselig	hostile	0,609	
verärgert	upset	0,532	0,110
begeistert	enthusiastic		0,781
entschlossen	determined		0,694
angeregt	inspired		0,682
stark	strong		0,665
aufmerksam	attentive	-0,158	0,630
stolz	proud	0,137	0,621
freudig erregt	excited	0,131	0,616
aktiv	active		0,576
interessiert	interested		0,569
wach	alert	-0,126	0,546

The pattern corresponds exactly to expectations. All items load on the appropriate factor and the secondary loadings are negligible. Although the Oblimin rotation was used, the correlation between the factors is.01. There is therefore not the slightest doubt about the orthogonality of the factors and the perfect mapping of the items.

Reliability of the PANAS scales

A reliability analysis was carried out for both scales. Table 6 shows the goodness of fit criteria Cronbach's α, McDonald's ω and the range of the corrected item total correlations.

Table 6. *Reliability. Cronbach's α, McDonald's ω, corrected item total correlation.*

	α	ω	corr. item total
PA	.87	.88	.51 –.72
NA	.88	.89	.51 –.71

Both scales have excellent internal consistency and neither could be improved by removing individual items. In view of the extraordinarily large sample, the corrected item total correlation can also be regarded as excellent.

Descriptive statistics of the dependent variables

The PANAS is a highly reliable instrument for Positive and Negative Affects. In addition, we consider the GM scale, which only consists of the two items „in a good mood“ and „cheerful“, which correlate with each other at.87 which gives a reliability of.93. The correlation with PA is.61; both scales measure positive emotional states, but the common variance is only 38 percent. The correlation with NA is -.33. Table 7 shows statistical parameters of the three scales.

Table 7. *Mean, standard deviation, skewness, kurtosis.*

	M	s	Skewness	Kurtosis
PA	1,80	0,79	0,19	-0,43
NA	0,57	0,66	1,74	3,29
GL	2,30	1,13	-0,24	-0,72

The exceptional role of Negative Affects immediately catches the eye. The mean value is only 0.57. Given a range of 0 to 4, this is a strong indication that the distribution is heavily skewed. This is confirmed on the one hand by the skewness index, and on the other hand we illustrate this in Figure 1.

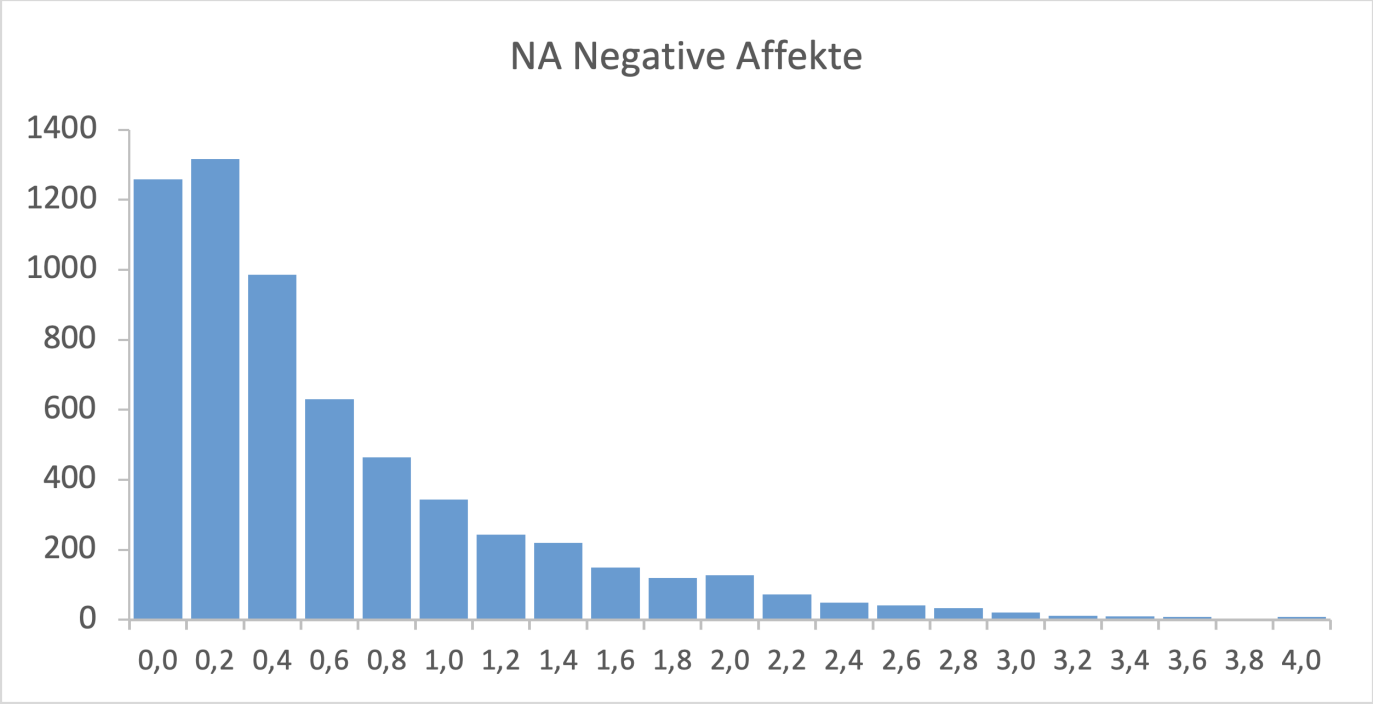

Figure 1. Frequency distribution. NA Negative Affects.

The distribution has nothing to do with a normal distribution. Obviously, the experimental setting generated virtually no negative feelings – which of course would not have been the intention. 1,258 participants even selected the value 0 on all 10 items and thus received the NA score 0. As the response pattern on the PA and GM scales is much more balanced, it is not reasonable to assume that these judges wanted to sabotage the study. Nevertheless, the ability to differentiate is severely limited by the extreme concentration at the lower end of the scale.

Effects of the independent variables on mood

The three mood indicators PA, NA and GM were subjected to an analysis of variance in a Color x Attractiveness x Sex x Language design. As there are significant age differences for sex and language version, in the first step age was taken into account as a covariate in an analysis of covariance (ANCOVA). After this, an ANOVA was carried out. Table 8 shows the p-values of all effects that are significant at the 5 percent level.

Table 8. *Significant effects, 5 per cent level.*

		PA	NA	GM
ANCOVA	Color	.018		.0345
	Attractiveness		.024	
	Language		<.001	.007
	F x A x G x S			.006
	Age	<.001	<.001	.002
ANOVA	Color	.016		.031
	Attractiveness		.034	
	Language		<.001	.010
	Sex	.009		
	F x A x G x S			.006

Table 8 shows an astonishingly simple picture. The design contains four two-way interactions, six three-way interactions and one four-way interaction. Of the 33 possible interaction effects, however, only a single one is significant, namely the four-way interaction on the GM scale. All others do not even reach the 10 percent level. As a four-way interaction cannot be meaningfully interpreted, only main effects remain.

The only difference between the ANCOVA and the ANOVA is that, without taking age into account, sex has a significant effect on the PA scale, which does not reach the 5 percent level in the covariance analysis (p =.075). Otherwise, the pattern is identical. In the following, we look at the results of the ANOVA.

For the *PA scale*, the Levene test is significant, i.e. the variance homogeneity requirement is not met ($F = 2.96$; $df1 = 47$; $df2 = 6070$; $p < .001$). However, the statistical significance is only due to the extraordinarily large sample. The Q-Q plot indicates an almost perfect normal distribution of the standardized residuals.

Positive Affects are significantly higher in men than in women. The difference is 0.06 points. This corresponds to an effect size of 0.07. However, when controlling for the age of the participants, the difference is no longer significant.

To shed light on the nature of the significant effect of color, pairwise post-hoc comparisons were carried out applying Tukey's and Holm's correction. According to both methods, only the difference between Green and Blue is significant at the 5 percent level. Positive Affects are 0.12 points higher with a green background, the effect size is 0.15. Green also tends to perform better than yellow (p_{Tukey} =.06; p_{Holm} =.08). The difference of 0.10 points corresponds to an effect size of 0.13.

For the *NA scale*, the significant Levene test (F = 4.21) is also of no importance. Here, however, the Q-Q plot indicates that the standardized residuals deviate strongly from the normal distribution. Given the extreme skewness of this scale, this is not surprising.

The Negative Affects are 0.04 points higher in the raters of the average to less attractive stimulus persons than in the raters of the extraordinarily attractive women. The effect size is 0.06.

There is a markedly greater difference between the language versions. In the English version, the mean value is 0.64, in the German version only 0.47. Thus, the skewness in the German version is much greater than in the English version (1.92 vs. 1.59). The effect size is 0.25.

For the *GM scale*, homogeneity of variance is fulfilled and the standardized residuals deviate from the normal distribution only at the extremes. Here too, the scores are higher in the English-language version. The difference of 0.08 points corresponds to an effect size of 0.07.

Just as with the Positive Affects, the post-hoc comparison Green – Blue is also significant on the GM scale. The difference is 0.15 points, the effect size is 0.13. Green – Yellow only narrowly fails the 5 percent mark and the parameters are also 0.15 points, d = 0.13. Green – Red also tends toward significance (0.14 points, d = 0.13, p_{Tukey} = 0.07; p_{Holm} = 0.09).

Overall, our two core variables have different effects. The attractiveness of the stimulus persons only had an impact on Negative Affects. However, the effect size of 0.06 is very weak. The background color of the website had an impact on both PA and GM. The differences between the color variants are illustrated in Figure 2 and Figure 3 showing the 95% confidence intervals.

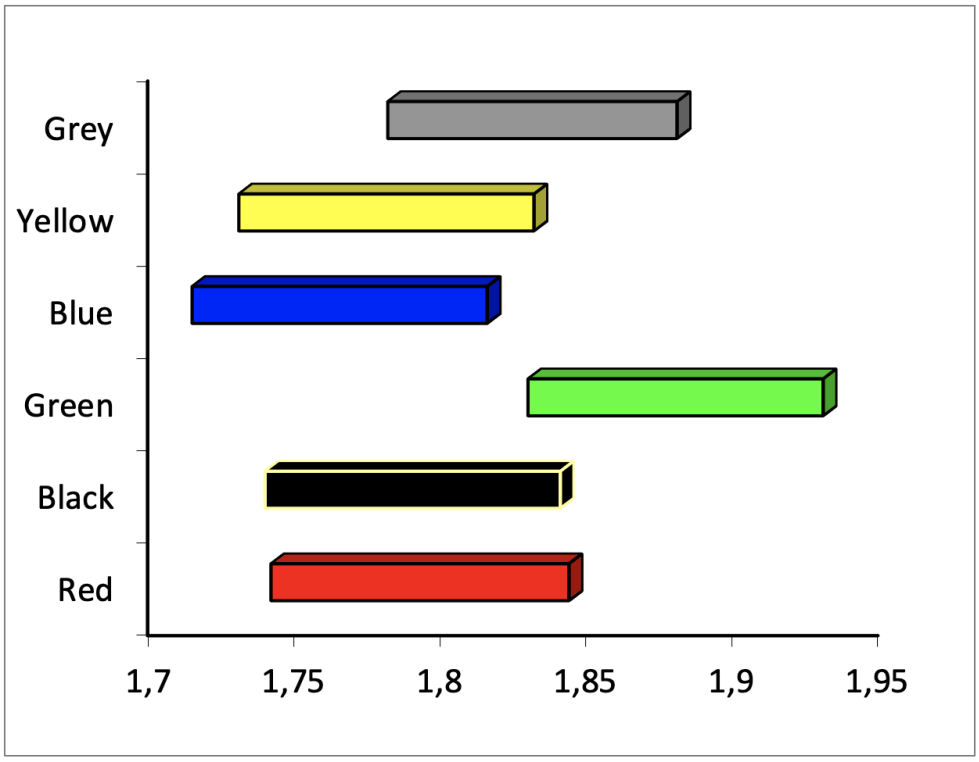

Figure 2. Background color and Positive Affects PA.

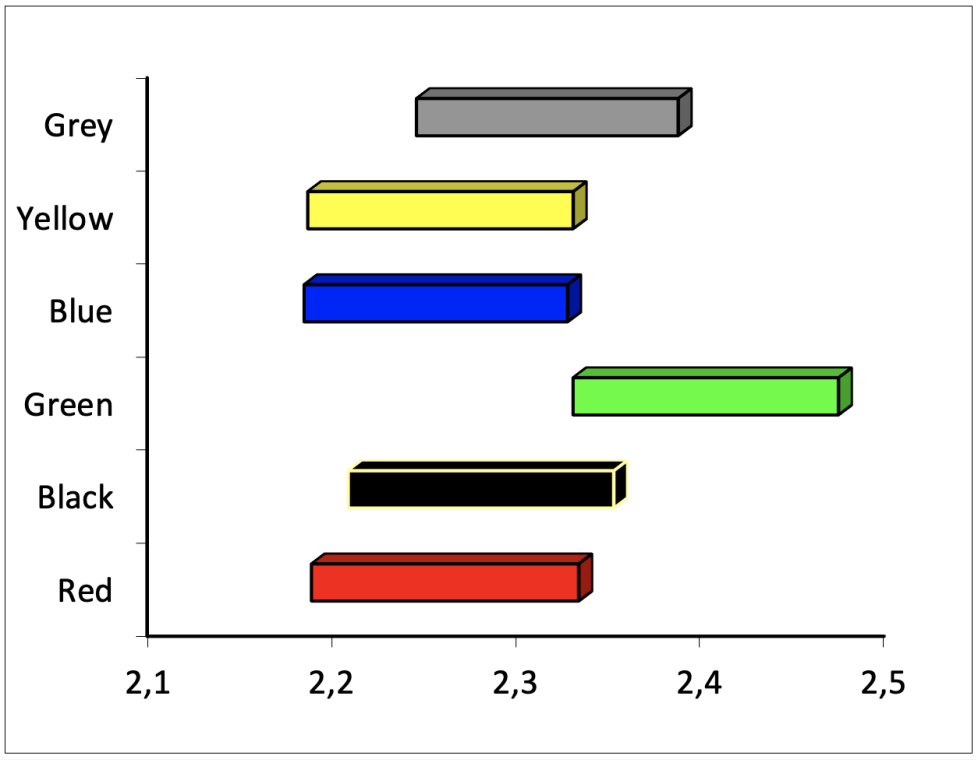

Figure 3. Background color and Good Mood GM.

Age and mood

Since, as seen in the covariance analysis (Table 8), age plays a significant role, we also look at the effect of age on

mood. Table 9 summarizes the mean values of the total sample for different age groups.

Table 9. *Age and current mood state.*

	14-19	20-24	25-29	30-34	35-39	40-44	45-74
PA	1,79	1,78	1,75	1,81	1,85	1,83	1,96
NA	0,68	0,59	0,54	0,48	0,38	0,44	0,45
GM	2,33	2,26	2,24	2,24	2,39	2,36	2,49

A one-way analysis of variance results in a p <.001 for each of the three scales. In all cases, the older participants perform „better“.

On the *PA scale*, the 45- to 74-year-olds scored significantly higher than the three youngest groups. The effect size is 0.21, 0.23 and 0.27.

On the *NA scale*, the under-20s scored significantly higher than all other groups (d = 0.14 to 0.46). The 20- to 24-year-olds reported significantly stronger Negative Affects than all groups aged 30 and over (d = 0.17 to 0.31). The 25- to 29-year-olds score significantly higher than the 35- to 39-year-olds (d = 0.24).

On the *GM scale*, the oldest group reported a more positive mood than the three groups aged 20 to 34 (d = 0.20; 0.22; 0.23).

Effects of color on the likeability ratings

Finally, we look at the impact of color on the likeability ratings. For each triple, the participants were asked to mark the most likeable one. The dependent variable is therefore the frequency with which the women were preferred over their two competitors.

Each of the two attractiveness levels comprises five triples. Thus, there are ten contingency tables with 3 (women) x 6 (colors) = 18 cells. The results of a chi-square test can be seen in Table 10.

Table 10. *p-values of Chi-square tests.*

	Triple				
Attractiveness	**1**	**2**	**3**	**4**	**5**
Very high	.004	.662	.680	.024	.303
Medium/Low	.290	.692	.015	.772	.140

Of the ten comparisons, only three are significant. In view of the extraordinarily large sample, however, only the first triplet of very attractive women is of interest (p =.004). First, however, we look at the fourth triplet of average to less attractive women in Figure 4, which is at the other extreme (p =.772). It is to be taken into account that in the medium/low group,

blue was seen somewhat more frequently than the other colors (533 vs. 496 to 507).

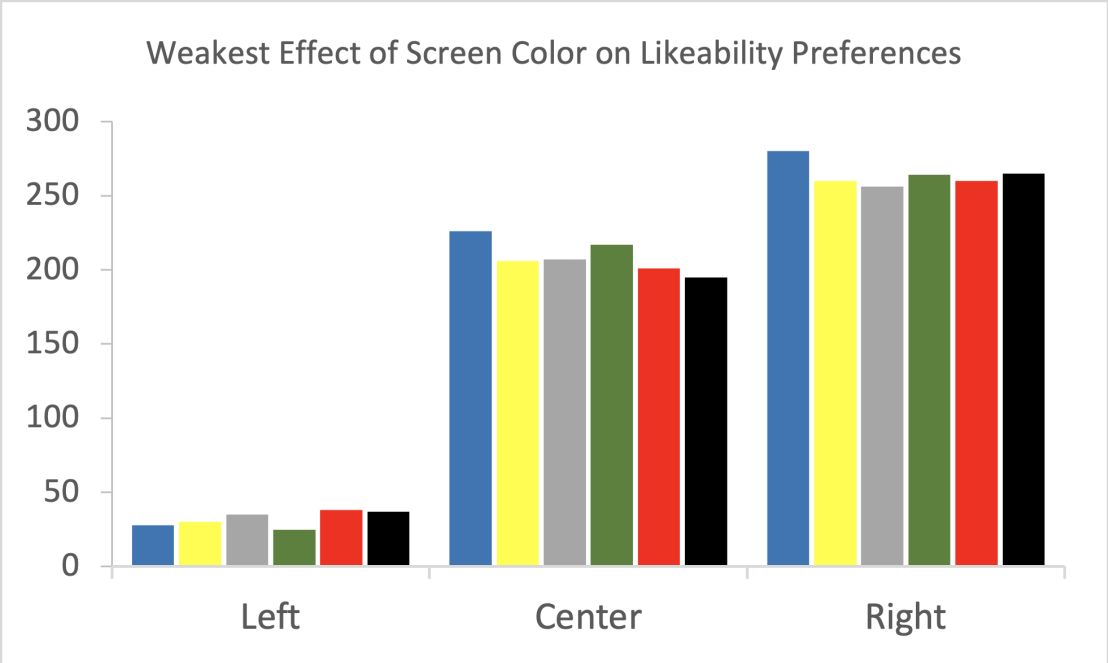

Figure 4. Weakest effect of screen color on likeability preferences.

The picture is unequivocal. The woman on the right is the most likeable and the woman on the left has the fewest votes. The ranking is the same for all color variants and the small deviations from this pattern are meaningless. The p-value of the chi-square test is.772, which means that one would expect larger differences when rolling the dice than is the case here.

Figure 5 shows the other extreme, i.e. the triple with the strongest effect of color. Here, gray was seen a bit more frequently (547 vs. 495 to 520).

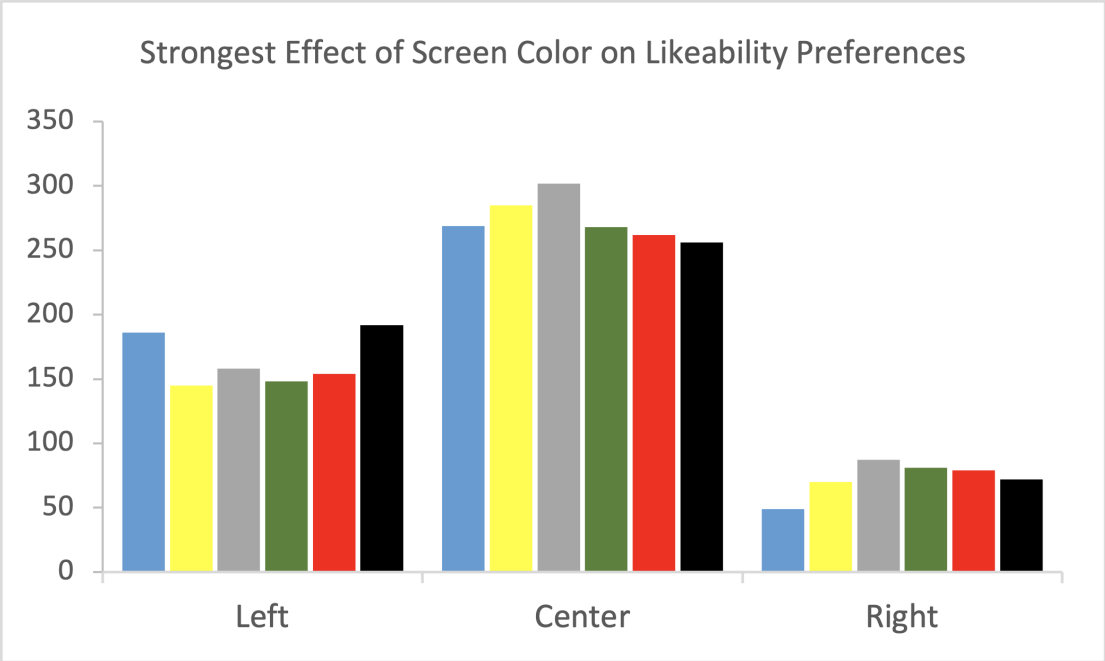

Figure 5. Strongest effect of color on likeability preferences.

In this triple, the effect of color is statistically significant (p =.004). The chi-square value is 25.7 (df = 10). Left/Blue contributes 3.6 points to this total, Left/Black 3.8 and Right/Blue 7.4. This means that with a blue background, the woman on the left gains some votes at the expense of the woman on the right. With a black background, she wins at the expense of the other two. The effect size, measured with Cramer's V, is 0.06. This is a very weak effect. [With Cramer's V, 0.1 is considered a weak effect, 0.3 a medium effect and 0.5 a strong effect.] In practical terms, the effect has no relevance at all. The ranking is Center, Left, Right and no matter which color variant one looks at, nowhere is there even a hint of a reversal of this order. A note is in order here: If there really is a neck-and-neck race, it may well happen that one woman is ahead in one color and the other in another color. In such a case, the shift may be statistically insignificant, but of enormous practical importance for those affected. In our case, this question does not arise; we have an indisputable pecking order and a statistically significant but very weak effect with no practical relevance.

Discussion

The most remarkable aspect of our study is the simplicity and clarity of the findings. This applies both to the measurement instruments and to the effects of the independent variables.

As there is a clearly defined factor structure for the PANAS, the first step was to carry out a confirmatory factor analysis. The model with two independent factors and uncorrelated errors shows a poor fit. By including correlated measurement errors, however, a very good fit can be achieved. This is well known in the field. An exploratory factor analysis employing the Maximum Likelihood method yields a perfectly orthogonal structure without noteworthy secondary loadings, even though an oblique rotation was applied. In addition, the reliability analysis attested excellent internal consistency to both scales.

The two factors are of different benefit for our research purposes. The NA scale is extremely skewed and thus the data are not suitable for most parametric analyses. However, the skewness is not a methodological shortcoming; it is in the nature of things. In a relaxed situation, in which only a few women's faces are to be judged according to their likeability, negative feelings evidently do not come into play at all. The clustering at the lower end is not an oddity of our setting. For example, Watson, Clark and Tellegen (1988) report a value of 0.74 for habitual mood (here and in all subsequent cases converted to a [0 – 4] scale). Breyer and Bluemke (2016) report 0.72 for a German sample; Krohne et al. (1996), also for a German sample, 0.84 for habitual mood and 0.47 for current mood. Coincidentally, the value of 0.47 corresponds exactly to ours. In a study on the social perception of genetically caused hair loss in men (androgenetic alopecia), in which photographs of men were to be assessed, we found values of 0.57, 0.61 and 0.62 in three German samples (Henss, 2024a). In everyday situations, Negative Affects are usually very weak, but there are of course numerous situations in which they are of great importance. In clinical studies in particular, Negative Affects are often stronger than Positive Affects (e.g. Estévez-López et al, 2016; Hovmand et al. 2023). As an example, the study by Díaz-García et al. (2020) may be considered. In adults with depressive, anxiety, and adjustment disorders, an online version of the PANAS yielded PA scores of 1.05, 1.00 and 0.91 and NA scores of 2.12, 2.04 and 1.80 in the pre-test. In the post-test, the PA scores were 1.64, 2.20 and 1.36, and the NA scores were 1.10, 1.07 and 1.18. Thus, there were impressive improvements on both scales, but the NA scores were still considerably higher in the post-test than in our study.

Although the ability to differentiate is severely limited in our study due to the clumping at the lower end, two effects are significant. The impact of the attractiveness of the stimulus persons is hardly worth mentioning. In contrast, the difference between the German and English versions (0.47 vs. 0.64) deserves attention, and not just because of the 0.25 effect size. We have shown that the Positive and Negative Affects are perfectly orthogonal. In a separate analysis, the correlation in the German version is -.14, whereas in the English version it is +.09.[5] Within both groups, the deviation from orthogonality is not dramatic, but since it points in opposite directions, the difference between the language versions is accentuated and at the same time the opposite tendency in the overall sample results in almost perfect independence.[6] Whether the skewness of the NA scale plays a role here is an open question. It is also unclear to what extent the difference is due to response tendencies, genuine differences in emotional experience, translation artifacts or other causes. Translations from one language to another are never perfect (Flores-Kanter et al., 2021; Torres et al., 2013; Wedderhoff et al., 2021) and it is not surprising that there are differences between the culturally and ethnically very homogeneous German-speaking sample and the heterogeneous international sample.

The PA scale shows one significant effect, namely for the background color of the website. The uncontested winner is Green. The 95 percent confidence interval shows no overlap with Blue and Yellow, with Red and Black the overlap is very small and there is also an advantage over Gray. Since the GM scale shows a very similar pattern, even though it covers slightly different aspects, there is some evidence that Green may actually lift people's mood. From everyday experience with green in nature, this seems plausible. However, such far-reaching associations cannot be derived from our data. We will come back to this at the end of this article.

In addition to color, the language version also has a significant effect on the GM scale. As with the NA scale, the value is

higher in the English-language version, but the effect size is considerably lower. On the PA scale, on the other hand, the values are slightly higher in the German version. This speaks against a difference in response sets.

Before turning to the two core variables of color and attractiveness, the age of the participants deserves attention. It is here that the strongest effects can be seen. Both in the positive and negative realms, older people perform better than young people. This fits in with the findings „that negative affect, as measured by the Positive and Negative Affect Schedule, decreases across the lifespan, whereas positive affect increases slightly until midlife and then decreases“ (Hoehne and Zimprich, 2024, p. 1; see also for example Flores-Kanter et al., 2021; Windsor and Anstey, 2010).[7] In online studies, the young are typically heavily overrepresented and the older ones heavily underrepresented and typically women predominate among the young and men among the older ones. In the age group [15 – 24] there are 44.8 percent of the men and 65.9 percent of the women, in the age group [40 – 74] the figures are 15.2 vs. 7.2 percent. Due to self-selection in online studies, sex and age are usually confounded and it is necessary to take age into account, even if, as in our case, it is not the focus of the research.

Our core variables are the background color of the website and the attractiveness of the stimulus persons. Attractiveness only had an impact on Negative Affects. The judges of the average to less attractive women show slightly higher values, but the effect is quite weak. It should be emphasized that the women were not particularly unattractive let alone ugly and that only five panels with three faces each were to be judged. Maybe judging a larger number of very unattractive and ugly faces would arouse negative emotions. Our study on male pattern baldness can also provide information here (Henss, 2024a). One sample rated the men on attractiveness, another on self-assurance and another on health. In terms of self-assurance and health, the ratings were somewhat above the centre of the scale, but the rating for attractiveness was extremely negative. However, the three samples showed very similar values for both Positive and Negative Affects. Remarkably, the sample that made the extremely low attractiveness judgements had slightly lower scores on the NA scale than the other two (0.57 vs. 0.61 and 0.62). In order to evoke negative emotions, one would probably have to present very ugly faces – but who would want to conduct such a study?

In retrospect, it is regrettable that only women’s faces were presented. It would have been better to use photos of men and women in two parallel versions. However, such an experiment would have been more difficult to conduct with male photos. The reason is that there are almost no exceptionally attractive men – at least not in the eyes of the participants in our online experiments. In our studies, we have repeatedly found that men perform worse than women in attractiveness assessments, often much worse (for example, Henss, 1992, 1998, 2024a, b). This is also found in many other studies. We recently published a particularly glaring case (Henss, 2024b). The stimulus material was photographs of 59 men taken from a model catalog. There is no doubt that men who offer their services through modeling agencies are, on average, much more attractive than their peers. The men were rated very favorably in terms of different personality traits, but the ratings on the goodlooking scale were below average, and in terms of sexual attractiveness they were demeaning. We refer to this phenomenon as the attractiveness malus for men. Whether photos of film stars and celebrities from show business and the world of sports would be more successful is an open question.

Our main interest is the effects of *color*. Two questions arise here: firstly, the effect on mood and secondly, the effect on

the likeability ratings.

For us, the effect on mood is of central interest, but in most online studies, the mood of the participants is not the focus of attention for the researchers and the clients. Whether the color scheme has an effect on the respective study's objective cannot be answered by our or any other study. In our case, the minimal effects on the likeability ratings are meaningless. However, this may be different in other experimental set-ups. There is no universal answer.

There is also no universal answer with regard to mood, but our study at least provides some information here.

The background color of the website has no influence on Negative Affects. The task of our study is probably pleasurable for most of the participants and in such a relaxed situation negative feelings do not come into play. The screen color is not expected to evoke negative emotions here. This is likely to be similar in many studies. However, there are also experimental set-ups that arouse strong negative emotions or in which the participants are already emotionally burdened. In these cases, the screen color might have effects that are demonstrable with a measuring instrument such as the PANAS.

The mood of our participants was balanced overall and there was sufficient variation with regard to positive emotions. There is a significant effect here, but it is rather weak. We deliberately chose intense colors in order to maximize possible effects. This may have been a misconception. It is quite conceivable that the intense colors inhibited positive emotions and that more discreet shades would increase the sense of well-being. This consideration leads to a very important point: Colors come in endless variations in terms of hue, saturation and brightness. We can only make statements about the specific variants that have been utilized. For example, Blue means the RGB value 000099 and Green 009900. However, we can give an answer to our question „Dark Screen – Dark Mood?". The answer is a firm No. On the NA scale, Blue performs no worse than the other colors, it does even a tiny bit better than Yellow, Grey and Black. And on the PA and GM scales, Blue is only outperformed by Green. This answer raises the question of the extent to which finding our „Green Screen – Good Mood" can be generalized.

The general question of the impact of screen color on mood has certainly been researched intensively in fields such as color psychology, emotion psychology, web design, human-computer interaction, the film industry, advertising, marketing, consumer behavior or art. We do not know to what extent the PANAS has been used in such studies. We also do not know whether green performs well in so many different situations that it can be called a mood enhancer and for which variants of green this applies.

We emphasized in the introduction that our study is exploratory and we have not formulated any hypotheses. In particular, we did not speculate about effects of certain colors. The reason is quite simple.

People like blue skies and the blue sea and blue is the color of loyalty. People like green leaves and meadows and green is the color of hope. They like red roses and red is the color of love. Yellow is reminiscent of the sun and gold and yellow is the color of happiness. Gray is November and dreariness. Black is night and death. However, there are also opposite associations for each color. Red is the blood and red signals danger. One turns red with anger, shame and agitation. Green stands for immaturity and poison, illness, envy, anger, disgust and jealousy. Yellow is a sign of envy, naivety,

stinginess, selfishness and infidelity. Blue appears cool and distant and in the English-speaking world it is considered the color of sadness. Grey is the transition between the light of day and the dark of night and appears harmonizing, calming, balancing and modest. There are even some positive associations for black, for example, black is considered the color of boldness and confidence. It stands for timeless elegance, practicality and functionality. The list could be continued at will and the contradictions become even more pronounced when different cultures are taken into account. Explanations that are based purely on color symbolism are little more than just-so-stories. Useful explanations must be sought at other levels.

Conclusion

The PANAS has excellent psychometric quality, but in a relaxed experimental set-up like ours there is an extreme concentration of negative affects at the lower end of the scale. The attractiveness of the women being assessed has a significant but weak effect on negative mood, which has no practical importance. Screen color has a significant impact on positive emotions, but the effect strength is weak. The screen color has no significant effect on the evaluation of women's likeability. The answer to the question „Blue Screen – Blue Mood?" is a firm no. The extent to which our finding „Green Screen – Good Mood" can be generalized is an open question.

_____

The data set ColoMoodDat.omv is freely available at https://www.researchgate.net/publication/378822387_ColoMoodDat

Other formats are available on request.

Footnotes

[1] We had only analyzed the data for internal use at the time and results were not published anywhere. We have now analyzed the data set in detail and believe the findings are still worth reporting.

[2] In other studies we also used the Mood Survey by Underwood and Froming (1980); Henss (1999).

[3] The panels were compiled in collaboration with 4 psychology students. People do not all have the same tastes, but when it comes to assessing attractiveness, the commonalities are much greater than many believe (Henss, 1992, 1995, 1998a; Todorov, 2017). There is certainly overlap in neighboring triplets, but very few would confuse the first and last triplets. The crucial point of our study, however, is the difference between the exceptionally attractive women and the average to unattractive ones. Only a vanishingly small minority, if anyone, would confuse them.

[4] This model has an excellent fit. However, it is purely data-driven and contains a dozen correlated error terms. An acceptable solution is also obtainable with fewer correlated measurement errors. However, a model should not be data-driven, but theory-driven. This task goes far beyond the scope of this paper.

[5] In the aforementioned study on the impact of hair loss (Henss, 2024a), we also found a slightly negative correlation between Positive and Negative Affects in three large samples from German-speaking countries, namely -10, -.14 and -.16 (Henss, 2024a). Breyer and Bluemke (2016) also report a slightly negative correlation of -.19 for a German sample. Schmukle, Egloff and Burns (2002) report a slightly negative correlation for an American sample at the state level, but not at the trait level.

[6] Our data set offers numerous possibilities for testing measurement invariance. But a systematic analysis would be far beyond our current issue.

[7] It should be noted that our study looks at the current mood, whereas other studies often look at the trait perspective.

References

- Alexander (2023). *The neuroscience of facial attractiveness*. https://datepsychology.com/the-neuroscience-of-facial-attractiveness/
- Breyer, B. & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel)* GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.6102/zis242
- Díaz-García, A., González-Robles, A., Mor, S., Mira, A., Quero, S., García-Palacios, A., Baños, R.M., Botella, C. (2020). Positive and Negative Affect Schedule (PANAS): psychometric properties of the online Spanish version in a clinical sample with emotional disorders. *BMC Psychiatry* (2020) 20:56. https://doi.org/10.1186/s12888-020-2472-1
- Estévez-López, F., Pulido-Martos, M., Armitage, C.J., Wearden, A., Álvarez-Gallardo, I.C., Arrayás-Grajera, M.J., Girela-Rejón, M.J., Carbonell-Baeza, A., Aparicio, V.A., Geenen, R., Delgado-Fernández, M. & Segura-Jiménez, V. (2016). Factor structure of the Positive and Negative Affect Schedule (PANAS) in adult women with fibromyalgia from Southern Spain: the al-Andalus project. *PeerJ* 4:e1822; https://doi.org/10.7717/peerj.1822
- Flores-Kanter, P.E., Garrido, L.E., Moretti, L.S. & Medrano, L.A. (2021). A modern network approach to revisiting the Positive and Negative Affective Schedule (PANAS) construct validity. *Journal of Clinical Psychology*, 77, 2370-2404. https://doi.org/10.1002/jclp.23191
- Henss, R. (1991). Perceiving age and attractiveness in facial photographs. *Journal of Applied Social Psychology*, 21, 933-946.
- Henss, R. (1992). *„Spieglein, Spieglein an der Wand …" Geschlecht, Alter und physische Attraktivität* Weinheim: Psychologie Verlags Union.
- Henss, R. (1995). Misses Germany, Titelmädchen des Jahres, Playmate des Jahres usw. Zur Urteilerübereinstimmung bei öffentlichen Schönheitswettbewerben und in psychologischen Untersuchungen. *Psychologische Beiträge*, 37, 124-139.
- Henss, R. (1997). *The Big Five and beyond*. Poster presented at the VIIIth. Biennial Meeting of the International Society for the Study of Individual Differences. At: Aarhus, Denmark, July 19-23, 1997. https://www.researchgate.net/publication/337312566_The_Big_Five_and_beyond
- Henss, R. (1998a). *Gesicht und Persönlichkeitseindruck*. Göttingen: Hogrefe. [Reprint: Henss, R. (2020). *Gesicht und*

Persönlichkeitseindruck. Saarbrücken: Dr. Ronald Henss Verlag. Freely available as PDF 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. (1999). Personality and Mood. A German version of the 'Mood Survey'. *Sportonomics*, 5, 73-76. https://www.researchgate.net/publication/337006799_Personality_and_Mood_A_German_version_of_the_'Mood_Survey'
- Henss, R. (2024a). „Toupee or not Toupee?" Sic (et Non). Social perceptions of Male Pattern Baldness (Androgenetic Alopecia). *Qeios*, https://doi.org/10.32388/UYKBF4
- Henss, R. (2024b). Hair loss, body height and the attractiveness malus for men. *Qeios*, https://doi.org/10.32388/NIAXZ7
- Hoehne, S. & Zimprich, D. (2024). Age-related differences in trait affect: Establishing measurement invariance of the Positive and Negative Affect Schedule (PANAS). *Psychology and Aging*. Advance online publication. https://dx.doi.org/10.1037/pag0000810
- Hovmand, O.R., Reinholt, N., Christensen, A.B., Eskildsen, A., Bach, B., Arendt, M., Poulsen, S., Hvenegaard, M. & Arnfred, S.M. (2023). Affectivity in Danish patients with emotional disorders: assessing the validity of the Positive and Negative Affect Schedule (PANAS). *BMC Psychiatry*, 23:943. https://doi.org/10.1186/s12888-023-05450-z
- Krohne, H.W., Egloff, B., Kohlmann, C. & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der 'Positive and Negative Affect Schedule' (PANAS). *Diagnostica*, 42, 139-156.
- Langlois, J.H., Ritter, J.M., Roggman, L.A., & Vaughn, L.S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology*, 27, 79-84.
- Langlois, J.H., Roggman, L.A., Casey, R.J., Ritter, J.M., Rieser-Danner, L.A. & Yenkins, V.Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype? *Developmental Psychology*, 23, 363-369.
- Liang, X., Zebrowitz, L.A. & Zhang, Y. (2010). Neural activation in the „reward circuit" shows a nonlinear response to facial attractiveness, *Social Neuroscience*, 5 (3), 320-334, https://doi.org/10.1080/17470911003619916
- Principe, C.P. & Langlois, J.H. (2011). Children and adults use attractiveness as a social cue in real people and avatars. *Journal of Experimental Child Psychology*, 115 (3), 590-597, https://doi.org/10.1016/j.jecp.2012.12.002
- Rhodes, G. & Zebrowitz, L.A. (2002). *Facial attractiveness. Evolutionary, cognitive, and social perspectives*. Westport, Connecticut: Ablex.
- Schmukle, S.C., Egloff, B. & Burns, L.R. (2002). The relationship between positive and negative affect in the Positive and Negative Affect Schedule. *Journal of Research in Personality*, 36, 463-475.
- Todorov, A. (2017). *Face value. The irresistible influence of first impressions*. Princeton, NJ: Princeton University Press.
- Torres, E.N.L., Silvestrin, M., Felix, K.S.A. & Biazoli, E.C. (2013). A systematic review of cultural adaptations of the Positive Affect and Negative Affect Schedule (PANAS). *PsyArXiv*, https://osf.io/preprints/psyarxiv/yx4s8
- Underwood, B. & Froming, W.J. (1980). The Mood Survey: A personality measure of happy and sad moods. *Journal of Personality Assessment*, 44, 404-414.
- 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.

- Wedderhoff, N., Gnambs, T., Wedderhoff, O., Burgard, T. & Bosnjak, M. (2021). On the Structure of Affect: A meta-analytic investigation of the dimensionality and the cross-national applicability of the Positive and Negative Affect Schedule (PANAS). *Zeitschrift für Psychologie*. https://doi.org/10.1027/2151-2604/a000434
- Windsor, T.D. & Anstey, K.J. (2010). Age differences in psychosocial predictors of positive and negative affect: A longitudinal investigation of young, midlife, and older adults. *Psychology and Aging*, 25 (3), 641-652. https://doi.org/10.1037/a0019431
- 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.