Check for updates

Stimulus Evaluation in the Eye of the Beholder: Big Five Personality Traits Explain Variance in Normed Picture Sets

Moritz Ingendahl[1] (ID), Tobias Vogel[2] (ID)

[1] *Department of Consumer and Economic Psychology, University of Mannheim, Mannheim, Germany.* **[2]** *Department of Social Sciences, University of Applied Sciences Darmstadt, Darmstadt, Germany.*

Abstract

The use of normed picture sets has become the gold standard in the study of affect, emotion, or attitudes. However, normed picture sets not only show the intended variance between pictures, but for each picture, normed ratings also show substantial variance between persons. Here, we examine whether interindividual variance in the pictures' evaluations is systematic and associated with personality traits. In a large-scale preregistered study, a heterogeneous sample of English- and German-speaking participants (total N = 901) completed a Big Five questionnaire and evaluated pictures of positive, neutral, and negative average valence from the OASIS database. The findings show that self-reported Neuroticism, Extraversion, and Agreeableness are associated with individual differences in picture evaluations, which supports and extends previous theorizing on personality and affect. Our results suggest that individual differences observed in paradigms employing valenced pictures may come from individual differences in picture evaluations rather than the processes under study.

Keywords

valence, Big Five, affect, attitudes

> **Relevance Statement**
>
> Standardized materials are indispensable for high-quality research in terms of objectivity, reliability, and validity. Thus, they are an essential ingredient for successful replication in psychology. In the last decades, psychological research has started using normed materials in a plethora of paradigms across disciplines. Yet, research on the validity of the materials themselves is rather scarce. Norms need to be tested regarding their validity across individuals to avoid systematic biases in analysis and interpretation. Here, we assess the person dependence of pictorial stimuli used in disciplines as diverse as emotion, neuroscience, social psychology, and cognitive psychology. For an illustration, we take the established taxonomy of the Big Five traits and a standardized database of valenced pictures. Our results indeed reveal that picture evaluations depend on personality traits. At the same time, our results emphasize the Big Five's role in understanding differences in emotional experiences.
>
> **Key Insights**
>
> - Big Five self-reports explain variance in standardized sets of affective pictures.
> - Neuroticism predicts more negative evaluations of negative pictures.
> - Extraversion predicts more positive evaluations of positive pictures.
> - Agreeableness predicts stronger effects of normed valence on evaluations.

Standardized materials are a key element in science and essential for achieving reliable, replicable, and internally valid findings. Whereas other scientific disciplines have been using standardized materials for a long time, psychology is still lagging behind (Lang & Bradley, 2007). To equip psychologists with standardized stimulus sets, several research groups assembled databases of stimuli, such as the International Affective Picture System (IAPS; Lang et al., 2008) or the Open Affective Standardized Image Set (OASIS; Kurdi et al., 2017). As an example, the OASIS consists of 900 license-free photographs with standardized information on their valence and arousal. Based on this information, the pictures are used as stimuli in research across disciplines, ranging from emotion research over psychopathology and neuroscience (Lang & Bradley, 2007) to social and cognitive psychology. Typical research paradigms that use valenced pictures are priming (Herring et al., 2013), attitude formation (e.g., Vogel et al., 2019), and attitude measurement (e.g., Kurdi & Banaji, 2019).

Despite clear norms for the valence of these pictures, a closer look also reveals substantial heterogeneity in the pictures' evaluations. As an example, picture I185 (showing a couple sitting on a bench) in the OASIS database yields a mean rating of 4.01, a seemingly neutral evaluation on the 1–7 scale. However, the standard deviation of the picture's valence ratings is 1.41. Assuming a normal distribution of the valence ratings, approximately 32% of a random participant sample would evaluate it as negative or positive (below 2.6 or above 5.4 on the 1–7 scale). In the OASIS database, the average

PsychOpen GOLD

standard deviation of a picture's valence rating is 1.10, indicating substantial heterogeneity in the evaluation of most pictures. Similar heterogeneity can also be found in other standardized sets (e.g., Lang et al., 2008). Where might this heterogeneity come from?

Interindividual Differences in Picture Evaluations

Evaluating pictures is a matter of complex appraisals and therefore prone to be influenced by many factors. Certainly, measurement error from single-item ratings used for most stimulus sets contributes to the large standard deviations. Even more systematic problems could be occurring with the response scale, as neutral ratings often represent ambivalent and not neutral attitudes (Schneider et al., 2016). However, in addition to situational influences, heterogeneity in picture evaluations could also reflect stable interindividual differences in psychological constructs. Previous research has already shown differences in valenced picture evaluations associated with sociodemographic variables such as age (Grühn & Scheibe, 2008) or gender (Lang & Bradley, 2007). Hence, differences associated with personality are also likely.

Notably, personality differences in the pictures' evaluations would have crucial implications for interpreting previous results obtained in different research paradigms using them. That is, variance in these paradigms might actually be explained by interindividual differences in picture evaluations rather than interindividual differences in the processes under investigation. For example, consider interindividual differences in evaluative conditioning, which refers to a change in stimulus liking because of its paired presentation with a positive/negative stimulus (Hofmann et al., 2010). Vogel et al. (2019) presented participants with conditioned stimuli (e.g., faces) together with either positive or negative pictures from a standardized set. They also assessed the Big Five personality traits (John & Srivastava, 1999) and found that evaluative conditioning was stronger for people high in Neuroticism or Agreeableness. While this might indicate that a person high in Neuroticism or Agreeableness is more likely to form associations between a neutral and a positive/negative stimulus, alternatively, it could mean that the pictures evoked more intense evaluations. This example makes clear that interindividual differences in picture evaluations are important to consider beyond personality research as they could change the interpretation of many findings.

To shed more light on interindividual differences in picture evaluations from standardized sets, we set out to study them in relation to the most prominent and accepted taxonomy of personality traits, the Big Five (John & Srivastava, 1999). In the next section, we introduce the Big Five and propose how they should relate to evaluations of valenced pictures.

Picture Evaluation and Big Five Personality Traits

In classic definitions, personality is the "coherent patterning of affect, behavior, cognition, and desires" (Revelle & Scherer, 2009, p. 304). Thus, many personality traits are associated with or even defined by the differential experience of positive/negative stimuli (Augustine & Larsen, 2015). Arguably, this is particularly true for the Big Five. The Big Five include the dimensions Openness (to experience), Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Out of these traits, Neuroticism and Extraversion are theoretically and empirically the most promising for this research question:

As noted by Costa and McCrae (1980, p. 673), "Extraversion [...] predisposes individuals toward positive affect, whereas Neuroticism [...] predisposes individuals toward negative affect". This view on Neuroticism and Extraversion is also reflected in classic personality theories (H. J. Eysenck & Eysenck, 1985; Gray, 1981) and supported by a plethora of empirical findings (Augustine & Larsen, 2015). However, it is less clear how this predisposition translates into behavior when evaluating pictures.

From an affect-level view, one should generally expect more positive affect for people high in Extraversion and more negative affect for people high in Neuroticism (Howell & Rodzon, 2011; Lucas & Baird, 2004). This perspective has received some empirical support (e.g., Gross et al., 1998; Howell & Rodzon, 2011; Lucas & Baird, 2004) and would predict a negative association of Neuroticism and a positive association of Extraversion with picture evaluations, irrespective of the pictures' valence.

From an affect-reactivity view, one would expect stronger reactivity of highly extraverted individuals to positive stimuli and highly neurotic individuals to negative stimuli. This perspective has also received empirical support (e.g., Canli et al., 2001; Gross et al., 1998; Larsen & Ketelaar, 1991; Rusting & Larsen, 1997; Smillie et al., 2012) and predicts more positive evaluations exclusively of positive pictures for individuals with higher Extraversion and more negative evaluations exclusively of negative pictures for individuals with higher Neuroticism. As both views agree on the latter associations, we conservatively expect:

> H1: Higher levels of Neuroticism are associated with more negative evaluations of negative pictures.

> H2: Higher levels of Extraversion are associated with more positive evaluations of positive pictures.

Regarding Agreeableness, there is less direct evidence of how it should be associated with picture evaluations. Yet, a vast amount of research has shown an overlap between disagreeableness and psychopathy (Decuyper et al., 2009; Stead & Fekken, 2014). People with high psychopathy show deviant reactions to emotional stimuli (Hoff et al., 2009; Kiehl et al., 2001). Correspondingly, Czerwon et al. (2011) found stronger valence judgments for both positive and negative faces for people with higher Agreeableness.

PsychOpen GOLD

Also, Vogel et al. (2019) found stronger evaluative conditioning effects for people with higher Agreeableness. In addition, with increasing levels of Agreeableness, people show stronger approach reactions towards positive pictures and stronger avoidance reactions towards negative pictures (Bresin & Robinson, 2015; Finley et al., 2017). Thus, we expect:

> H3: Higher levels of Agreeableness are associated with more positive evaluations of positive pictures.

> H4: Higher levels of Agreeableness are associated with more negative evaluations of negative pictures.

Next to positive or negative pictures, personality might be the deciding factor whether a neutral picture is actually rather seen as positive or negative. Thus, some of the previously mentioned relationships could also be present for neutral pictures. Indeed, neutral ratings in picture evaluations often reflect mixed responses towards ambivalent pictures (Schneider et al., 2016). Neutral pictures could represent "weak situations" in which associations with personality are the strongest. People high in Neuroticism, as an example, are more likely to interpret even ordinary situations as threatening (e.g., Lommen et al., 2010). However, the theoretical basis for directional hypotheses on neutral pictures is much weaker than for positive or negative pictures. Therefore, we refrain from formulating explicit hypotheses here.

Lastly, for the remaining traits of Conscientiousness and Openness, there is considerably less theoretical or empirical background than for the other three traits to make assumptions about how they might be associated with picture evaluations (Augustine & Larsen, 2015). Thus, we want to examine the association of the two traits with picture evaluations in an exploratory manner.

Despite the relevance of the issue and a plethora of previous research on personality and affect in general (Augustine & Larsen, 2015), empirical evidence for the association of the Big Five and picture evaluations in standardized sets is scarce. Of the few studies that have touched on the question, one did not include neutral pictures and used a statistical model not suited to answer our research question (Tok et al., 2010). Another study did not employ the Big Five but Impulsiveness and Anxiety (Aluja et al., 2015) and thus offers limited insight for our purposes. More recently, Levine and colleagues (2020) investigated interindividual differences in how participants cluster pictures from the IAPS. Since this study did not assess any valence ratings, it is not directly applicable, but their results do suggest substantial interindividual differences in how participants cluster pictures depending on the Big Five. Thus, previous research attests to the importance of our research question but also shows that a (more) systematic investigation is necessary to answer it.

PsychOpen GOLD

Overview of the Present Research

In this research, we examine to what extent picture evaluations from standardized sets are associated with the Big Five. For that purpose, we administer the BFI-2 (Danner et al., 2019; Soto & John, 2017) and let participants evaluate pictures of different normed valence in the OASIS (Kurdi et al., 2017). As population estimates for correlations with personality traits require large and also heterogeneous samples (Schönbrodt & Perugini, 2013), we collected data from 936 German-speaking and English-speaking participants of different ages, gender, and education.

We preregistered our hypotheses, methods, and analyses on the OSF: https://doi.org/10.17605/OSF.IO/92WSU. All data, analysis scripts, and materials are provided in the Supplementary Materials.

Method

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Design and Participants

In a single-factor design, normed valence of the pictures (positive vs. neutral vs. negative) varied within participants. The Big Five served as continuous covariates. To determine the sample size, we conducted an a priori power analysis with GPower (Faul et al., 2007). As a rough approximation for our design and analysis, we took the mixed ANOVA design with two groups and three repeated measures (α = .05, 1-β = .9). Our goal was to detect a small effect size of f = .1 for the between-within interaction and the between-participants main effect, which resulted in N = 214 and N = 704 as a minimum sample size. We thus aimed at a minimum sample size of N = 800. To achieve the necessary power and trait heterogeneity, 936 participants were recruited via the Respondi panel. We considered only finished interviews. English-speaking participants (53.95%) came from the UK and German-speaking participants from Germany, Austria, and Switzerland. Participants were compensated according to the panel's incentive system (min. 1€) for a 20-minutes study that consisted of multiple independent tasks. Detailed descriptive statistics of our sample are displayed in Table 1. Overall, our sample was very heterogeneous regarding gender, age, and education.

Table 1

Sample Characteristics

	Language	
Characteristic	**English**	**German**
Total Sample	505	431
Gender		
Male	275	235
Female	229	194
Non-binary	1	2
Age		
M_{age} (*SD, Min, Max*)	53.3 (14.82, 18, 84)	49.49 (15.28, 18, 88)
Language Proficiency		
Native speakers	375	397
Fluent	105	25
Education Level		
University degree	177	119
A-level/Abitur	149	89
Middle school/Realschule	30	149
Secondary school/Hauptschule	119	47
Primary school/Grundschule	5	12
No formal education	8	0
Other	17	15

Note. Education levels of English- and German-speaking participants do not correspond exactly due to differences in the education systems. If not indicated otherwise, numbers represent frequencies.

Procedure and Materials

After providing informed consent, participants first filled out the personality questionnaire. Next, 30 pictures (10 per valence level) were drawn randomly via the PHP shuffle function from our stimulus pool and presented to the participants who rated them on valence. After this evaluation task, participants proceeded with other tasks unrelated to this research question[1], demographic information was assessed, and participants were thanked and debriefed about the study's purpose. In line with our university's ethics committee guidelines, the study did not require specific approval. We received approval from our university's data protection office.

---

1) Specifically, participants did further judgment tasks (e.g., evaluating fictional letter strings). As these data have not been published (yet), we provide only the data that are relevant to our research question in the Supplementary Materials. All measures and variables relevant to our research question are reported in this manuscript.

Big Five Measures

For measuring the Big Five, we used the BFI-2 with 60 items[2] (German: Danner et al., 2019; English: Soto & John, 2017). Descriptive statistics, Cronbach's alphas, and intercorrelations are provided in Table 2.

Table 2

Intercorrelations, Cronbach's Alphas (Main Diagonal), and Descriptive Statistics of the Big Five Personality Traits (N = 901)

Personality Trait	N	E	A	C	O
N	(.92/.91)	-.37/-.38	-.31/-.34	-.41/-.44	-.07/-.23
E		(.81/.85)	.18/.27	.33/.28	.35/.38
A			(.83/.82)	.40/.32	.22/.22
C				(.89/.86)	.21/.21
O					(.82/.84)
M	2.73/2.69	3.13/3.19	3.78/3.72	3.82/3.74	3.46/3.38
SD	0.85/0.75	0.64/0.62	0.61/0.51	0.69/0.61	0.66/0.64

Note. For all items, the scale ranged from 1 to 5. The first value corresponds to the English-speaking participants, whereas the second value corresponds to the German-speaking participants. N = Neuroticism; E = Extraversion; A = Agreeableness; C = Conscientiousness; O = Openness. All correlations are significant at $p < .001$, except for the correlation of Neuroticism and Openness in the English-speaking sample. According to five exploratory *t*-tests, German- and English-speaking participants did not differ significantly in mean trait levels (all *p*'s > .069). Following the preregistration, we nevertheless standardized the Big Five within each language for all following analyses to avoid any potential confound by language. The intercorrelations and internal consistencies were similar to those reported in the original publications (Danner et al., 2019; Soto & John, 2017).

Valenced Pictures

As variation of our experimental factor 'normed valence', we selected 3 x 30 pictures from the OASIS (Kurdi et al., 2017) as stimuli. We used the OASIS because the pictures are current and have high quality. Furthermore, the pictures are license-free and thus usable in online research. Thirty pictures with normed valence ratings higher than +1 *SD* above the mean of all OASIS pictures were chosen as positive stimuli (OASIS valence rating > 5.56 on the scale of 1-7), thirty pictures with normed valence ratings between -0.33 *SD* (3.99) and +0.33 *SD* (4.71) as neutral stimuli, and 30 pictures with normed valence ratings below -1 *SD* (2.86) as negative stimuli. Arousal was kept constant at medium levels, with normed arousal ratings between -1 *SD* (2.86) and +1 *SD* (4.50). Hence, pictures obviously differed in their valence ratings reported in the manual, $F(2, 87) = 1559.17$, $p < .001$, $\eta^2 = .97$, 95% CI [.96, .98], $M_{neg} = 2.40$, $M_{neu} = 4.42$, $M_{pos} = 5.88$, but not in

2) Note that the BFI-2 labels Neuroticism as Negative Emotionality and Openness as Open-mindedness, but we use the terms more common in the literature here.

their arousal ratings, $F(2, 87) = 1.76$, $p = .178$, $\eta^2 = .04$, 95% CI [.00, .13]. Also, the sets had similar valence standard deviations reported in the manual, $F(2, 87) = 1.19$, $p = .308$, $\eta^2 = .03$, 95% CI [.00, .11], and the mean valence ratings within a set were similarly heterogeneous according to a Levene test, $F(2, 87) = 2.13$, $p = .125$. Each valence set consisted of ten images depicting scenes, ten depicting persons, five depicting objects, and five depicting animals. A list of the used stimuli is provided in the Supplementary Materials. We did not use pictures that depicted extreme violence or nudity.

For the evaluation task, we used the same instructions and rating scale format (a seven-point scale labeled at each point) as in the OASIS (Kurdi et al., 2017)[3]. For the German-speaking participants, all instructions and materials were first translated by the online tool deepl, and then translations were slightly modified by two German native speakers who are proficient in English (C2 level; instructions are provided in the Supplementary Materials). As in the original OASIS norming study, each picture was presented on a single slide. Below the picture, the heading "Valence" was presented together with the labeled scale (very negative to very positive). For the 90 pictures, the aggregated valence ratings in our study almost perfectly correlated with those reported in the OASIS manual, $r(88) = .98$, $p < .001$, also when using only the German-speaking participants.

Exclusion Criteria

Following the preregistration protocol, we excluded 35 participants who provided the same answer for over 25 pictures in the evaluation task, or the same answer for more than 50 items of the BFI-2, leading to a final sample of 901 participants. This exclusion criterion was chosen to exclude participants that answered redundantly (e.g., giving the same response to pass the study as fast as possible). In an exploratory manner, we also repeated the analyses without any exclusions. These analyses yielded nearly identical results and are thus only provided as an HTML document in the Supplementary Materials.

Results

Preregistered Analytical Approach

As measurements were nested within participants, we ran multilevel regression models with the R package lme4 (Bates et al., 2019). In a first baseline model, we decomposed the variance of the picture evaluations by including random intercepts for the partici-

---

3) Note that these instructions tell participants to rate the *pictures* and not the *feelings* the pictures evoke. However, in the development of the OASIS both picture- and feeling-focused instructions were pretested and led to the same ratings, suggesting that our findings should not depend on specific instructions (Kurdi et al., 2017).

pant and the specific OASIS picture. There was substantial variance between pictures, SD = 1.52, which is not surprising given that the pictures had been selected to capture a large range of valence, but there was also variance between participants, SD = 0.35, next to residual variance, SD = 1.21.

Next, we added two dummy variables for positive and negative normed picture valence in the model (see Model 1 in Table 3). We allowed these effects to vary between participants by including random slopes in this and all following models. As expected, positive pictures were evaluated more positively, b = 1.49, 95% CI [1.30, 1.68], and negative pictures more negatively, b = -2.09, 95% CI [-2.29, -1.80], than neutral pictures. Notably, the effects of positive and negative pictures were heterogeneous across participants, as indicated by the standard deviation of the random slopes, SD_{pos} = 0.29, SD_{neg} = 0.56.

Table 3

Picture Evaluations Predicted by Normed Valence and Big Five Personality Traits

	Model 1				Model 2				
		95% CI				95% CI			Estimate with
Model Term	Estimate	LL	UL	p	Estimate	LL	UL	p	control variables
Intercept	4.42	4.28	4.55	< .001	4.42	4.28	4.55	< .001	4.42
Positive	1.49	1.30	1.68	< .001	1.49	1.30	1.68	< .001	1.49
Negative	-2.09	-2.29	-1.90	< .001	-2.09	-2.29	-1.90	< .001	-2.09
N					-0.04	-0.08	0.01	.149	-0.01
N * Positive					0.10	0.04	0.15	< .001	0.05[a]
N * Valence					-0.08	-0.14	-0.01	.023	-0.05[a]
N (Positive)					0.06	0.01	0.11	.013	0.03[a]
N (Negative)					-0.11	-0.17	-0.06	< .001	-0.06
E					0.11	0.07	0.16	< .001	0.12
E * Positive					-0.00	-0.06	0.04	.890	-0.01
E * Negative					-0.13	-0.20	-0.07	< .001	-0.13
E (Positive)					0.11	0.06	0.16	<.001	0.11
E (Negative)					-0.02	-0.07	0.03	.452	-0.01
A					0.05	0.01	0.10	.030	0.07
A * Positive					0.20	0.15	0.25	< .001	0.16
A * Negative					-0.19	-0.25	-0.13	< .001	-0.19
A (Positive)					0.25	0.20	0.29	< .001	0.23
A (Negative)					-0.14	-0.19	-0.09	< .001	-0.12
C					0.00	-0.04	0.05	.809	0.02
C * Positive					0.06	0.00	0.11	.040	0.03[a,b]
C * Negative					-0.09	-0.15	-0.02	.008	-0.09
C (Positive)					0.06	0.01	0.11	.010	0.05
C (Negative)					-0.08	-0.14	-0.03	.003	-0.07

| Model Term | Model 1 | | | | Model 2 | | | | Estimate with control variables |
| | Estimate | 95% CI | | p | Estimate | 95% CI | | p | |
		LL	UL			LL	UL		
O					0.03	-0.02	0.07	.229	0.03
O * Positive					0.01	-0.04	0.06	.819	0.01
O * Negative					-0.05	-0.11	0.01	.115	-0.05
O (Positive)					0.03	-0.01	0.08	.139	0.04
O (Negative)					-0.02	-0.07	0.03	.413	-0.03
Random Effect *SD*s									
Participants	0.55	0.52	0.59		0.53	0.49	0.56		
Pictures	0.36	0.31	0.42		0.36	0.31	0.42		
Positive	0.54	0.49	0.59		0.50	0.45	0.54		
Negative	0.75	0.70	0.80		0.69	0.64	0.74		
Residual	1.11	1.10	1.12		1.11	1.10	1.12		

Note. All models were run with $N = 901$. The Big Five were standardized. Confidence Intervals were computed with the confint.merMod function of the R package lme4 (Bates et al., 2019) with the profile method. N = Neuroticism; E = Extraversion; A = Agreeableness; C = Conscientiousness; O = Openness. N * Positive refers to the interaction of Neuroticism and positive normed valence of a picture; N (Positive) refers to the simple effect of Neuroticism for pictures with positive normed valence. Model 2 was repeated with the control variables age, gender, and language, and also when aggregating within valence level and participants. Detailed results of these analyses are provided in the Supplementary Materials (Tables A1 and B1).
[a]Change in significance ($\alpha = .05$) in the control variable model.
[b]Change in significance ($\alpha = .05$) in the aggregated model.

For our main model, we standardized the Big Five within languages. They were entered together with their two-way interactions with the two dummy variables into the model (see Model 2 in Table 3). This reduced the variance between participants, $SD = 0.53$, and the variance of the slopes, $SD_{pos} = 0.25$, $SD_{neg} = 0.48$, but not the variance between the pictures, $SD = 0.13$. In this model, the main effect of a Big Five trait refers to the effect for neutral pictures. The interaction term of positive/negative valence with a Big Five trait can be interpreted in two ways: First, it captures the extent to which interindividual differences in valence effects (i.e., the variation in the random slopes) can be explained statistically by the respective trait. Second, it captures changes in the main effect when a picture is positive/negative. We reran this model with different valence dummy variables to obtain simple slope estimates for each valence level, thereby changing the baseline. Due to our coding scheme, regression weights can be interpreted as the increase on the 7-point scale by steps of 1 standard deviation on a personality trait. The results of these analyses are visualized in Figure 1.

Figure 1

Picture Evaluation by Normed Valence and Big Five Personality Traits
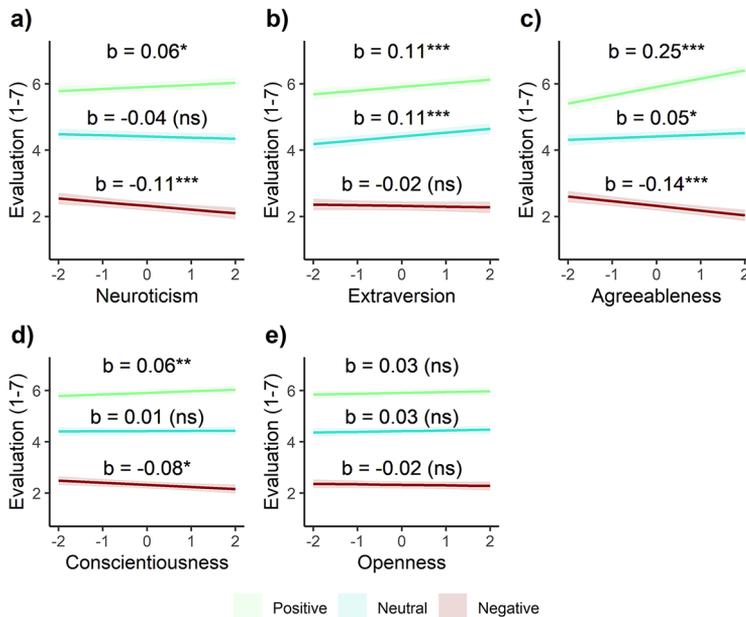
Note. Effect sizes refer to Model 2 in Table 3. The Big Five were standardized. Effect sizes can be interpreted as increase/decrease on the 1-7 rating scale for +/-1 *SD* on a Big Five trait. Shaded areas represent 95% confidence intervals.

p < .05. **p* < .01. ***p* < .001. ns = not significant.

Main Results

As can be seen in Figure 1 and Table 3, higher levels of Neuroticism were not associated with evaluations of neutral pictures, *b* = -0.04, 95% CI [-0.08, 0.01], but with more negative evaluations of negative pictures, *b* = -0.11, 95% CI [-0.17, -0.06]. This is consistent with Hypothesis 1. Unexpectedly, higher levels of Neuroticism were also related to more positive evaluations of positive pictures, *b* = 0.06, 95% CI [0.01, 0.11].

Higher levels of Extraversion were associated with more positive evaluations of positive pictures, *b* = 0.11, 95% CI [0.06, 0.16], therefore supporting Hypothesis 2. However, this positive association was also present for neutral pictures, *b* = 0.11, 95% CI [0.07, 0.16], but not for negative pictures, *b* = -0.02, 95% CI [-0.07, 0.03].

Consistent with Hypothesis 3 and 4, higher levels of Agreeableness were associated with more positive evaluations of positive pictures, *b* = 0.25, 95% CI [0.20, 0.29], and more negative evaluations of negative pictures, *b* = -0.14, 95% CI [-0.19, -0.09]. In addition,

higher levels of Agreeableness were weakly positively related to the evaluations of neutral pictures, b = 0.05, 95% CI [0.01, 0.10].

Next to the hypothesized relevant traits, we also observed associations of Conscientiousness with picture evaluations: Higher levels of Conscientiousness were associated with more positive evaluations of positive pictures, b = 0.06, 95% CI [0.01, 0.11] and more negative evaluations of negative pictures, b = -0.08, 95% CI [-0.14, -0.03]. Finally, Openness was not significantly associated with evaluations of neutral, positive, or negative pictures (see Table 3). Thus, overall, our hypotheses were supported by the data.

Robustness Checks

Following the preregistration protocol, we next examined to what extent the found associations between the Big Five and picture evaluations were incremental to associations with sociodemographic variables. Both age and gender have been found to correlate with picture evaluations (Grühn & Scheibe, 2008; Lang & Bradley, 2007) and the Big Five (e.g., Donnellan & Lucas, 2008). Therefore, we repeated Model 2 while controlling for gender (standardized, higher scores = male), age (standardized), and language (standardized, higher scores = English). All control variables were allowed to interact with the two valence dummies. We only report results here if the inclusion of the control variables changed the significance (p < .05) of one of the terms (see Table 3). Detailed results of these analyses are provided in the Supplementary Materials (Table A1). Overall, the associations postulated in our hypotheses remained stable when controlling for sociodemographics. Only the association of Neuroticism with evaluations of negative pictures (H1) was weaker but still significant, and the unexpected association of Neuroticism with evaluations of positive pictures was not significant anymore.

Due to a nonnormal distribution of residuals, we repeated Model 2 with aggregated data as a statistical robustness check (which was not preregistered). The ten measures per individual and valence level were aggregated, which led to a model that only contained three measures per participant and thus only intercepts of the participants as random effects. Again, the results did not change much (see Table 3). The detailed results of these analyses can be found in the Supplementary Materials (Table B1).

Exploratory Facet-Level Analyses

In an exploratory manner, we also repeated our main model with the 15 facets of the BFI-2 (instead of the five dimensions) as predictors. As these analyses were not preregistered, we do not present the results here but in the Supplementary Materials (Table E1). In essence, none of the Neuroticism facets were significantly associated with picture ratings. For all other dimensions, the facets showed diverging and sometimes even opposing associations (e.g., Productiveness was positively associated but Responsibility

negatively associated with ratings of negative pictures, despite both belonging to the dimension Conscientiousness).

Discussion

In this research, we examined whether the Big Five personality traits are associated with valence ratings of pictures from a standardized database. Our preregistered large-scale study (N = 901) revealed that all Big Five traits except Openness are associated with evaluations of positive, neutral, or negative pictures. In the following, we first discuss these results in light of personality research, then implications for research that uses valenced pictures, and finally limitations of our study.

Big Five Traits and Picture Evaluations

Overall, our predictions for Neuroticism and Extraversion were validated by the pattern of results: Based on classical personality theories (H. J. Eysenck & Eysenck, 1985; Gray, 1981), Neuroticism should be associated with negative and Extraversion with positive affect. Consistent with this, higher levels of Neuroticism were associated with more negative evaluations of negative but not positive pictures. Correspondingly, higher levels of Extraversion were associated with more positive evaluations of positive but not negative pictures. These results are consistent with the affect-reactivity perspective on the link between personality and affect (Canli et al., 2001; Gross et al., 1998; Larsen & Ketelaar, 1991; Rusting & Larsen, 1997; Smillie et al., 2012). However, there was also a positive relationship of Extraversion with the ratings of neutral pictures. This might also fit the affect-reactivity perspective, as previous research has shown that neutral evaluations often reflect ambivalent attitudes (Schneider et al., 2016). Highly extraverted individuals might react positively to the positive aspects of a picture. However, it is also possible that participants evaluated the pictures in a way consistent with how they *wanted* to feel. For instance, extraverts prefer experiencing positive emotions and might thus evaluate the neutral pictures consistent with their preferred emotional state (e.g., Tamir, 2009).

Our results regarding Agreeableness also matched with our theoretical reasoning – higher self-reported Agreeableness was associated with more positive evaluations of positive pictures and more negative evaluations of negative pictures. In other words, high Agreeableness emphasizes the valence implied in the norm ratings. This fits the idea that less agreeable individuals show deviant affective reactions to emotional stimuli (Decuyper et al., 2009; Stead & Fekken, 2014) and is in line with previous research on Agreeableness and emotional stimuli (Bresin & Robinson, 2015; Czerwon et al., 2011; Finley et al., 2017; Vogel et al., 2019). Apparently, people high in Agreeableness are most likely to show the consensual reaction, thus a positive (negative) evaluation of pictures that are evaluated positively (negatively) by the vast majority of people.

PsychOpen GOLD

However, there might also be alternative explanations of our findings on Agreeableness. For instance, previous research has shown that agreeable individuals are emotionally more responsive to social situations relevant to interpersonal relationships (Tobin et al., 2000). This would imply that the pronounced relationships of Agreeableness with picture evaluations might be due to pictures depicting social content. We thus reran our main model with an additional dummy variable that codes picture content (1 = social, 0 = non-social). These exploratory analyses revealed that the relationships of Agreeableness and picture evaluations were descriptively slightly stronger for the social pictures but still present for the non-social pictures. Detailed results of these analyses are provided in the Supplementary Materials (Table C1). As a further interpretation, agreeable participants might simply be more compliant with the evaluation task and thus provide more reliable ratings (Vogel et al., 2019). Overall, our findings show that Agreeableness and its role in affective processes deserve to be further examined in future research.

However, our results also revealed some effects that were not predicted by us, mainly regarding Conscientiousness. Higher levels of Conscientiousness were associated with more positive evaluations of positive pictures and more negative evaluations of neutral and negative pictures. One major aspect of Conscientiousness is acting dutifully and focused (Soto & John, 2017). Therefore, we speculate that people with high Conscientiousness simply took the task more seriously and provided more reliable judgments.

Overall, our results show that the Big Five are indeed related to interindividual differences in valence ratings from a standardized database, even beyond sociodemographic variables. This is consistent with previous research on personality and affect and further contributes to this field. Next to personality research, however, these results could also have important implications for psychological paradigms in other disciplines.

Implications Beyond Personality Research

We started this paper by arguing that systematic interindividual differences in picture evaluations could pose a potential problem for prominent paradigms in psychology, and our results indeed suggest that this may be the case. To pick up the introductory example, Vogel et al. (2019) found stronger evaluative conditioning effects for people with higher Neuroticism and Agreeableness – two traits for which we also found more pronounced effects of a picture's valence. Thus, it seems likely that those traits do not moderate the conditioning process itself, but the unconditioned stimuli have a stronger valence for people with higher Neuroticism and Agreeableness. Clearly, future research is necessary to examine the Big Five x Evaluative Conditioning moderations further.

This also raises the question of how researchers using these paradigms should deal with interindividual differences in the pictures' evaluation. On the one hand, researchers who want to avoid any association with personality could select only those pictures that are evaluated positively/negatively by the individual participant. Another possibility

would be to control for interindividual differences statistically (i.e., adding pre-ratings as covariates). On the other hand, detecting (instead of avoiding) associations with personality could also improve our knowledge on the underlying mental processes in these paradigms. For instance, the fact that participants high in Extraversion evaluate positive pictures more positively but apparently do not show elevated conditioning effects (Vogel et al., 2019) could imply that some of the processes underlying evaluative conditioning are weaker amongst extraverts.

Finally, we are confident that our results have similar implications for other research designs or even other research areas (e.g., neuroscience). On the positive side for research using these pictures, one should also keep in mind that we find modest effect sizes, with maximum shifts of 1/4 scale points on the 1–7 scale for a +/- 1 *SD* increase/decrease on a trait. Still, we focused exclusively on the conceptually broad Big Five in our research – more narrow personality traits (e.g., Need for Affect, Maio & Esses, 2001) could lead to even more pronounced effects. However, this is just a speculation, which brings us to the limitations of our research.

Limitations and Future Research Directions

Our findings are restricted to a selection of 90 pictures. We chose this stimulus pool for its similarity to those used in a typical psychological paradigm (such as evaluative conditioning) regarding size, symmetric differences in valence, but no differences in arousal. Future research should aim to replicate our findings with another selection of pictures, perhaps even from other popular standardized databases, such as the IAPS (Lang et al., 2008).

Also, we used the rating paradigm and instructions from the OASIS to make our findings comparable to the normed ratings. However, such ratings only capture rather spontaneous appraisals but do not assess temporal dynamics of affective reactions. In many research paradigms, the same picture is presented either for a longer time or on multiple occasions. Previous research has shown that the Big Five are also associated with emotion regulation processes (Augustine & Larsen, 2015; Bresin & Robinson, 2015). Thus, future research should investigate whether found associations of the Big Five and evaluations of affective pictures also vary over time.

In addition, we tested our hypotheses in a broad sample of German- and English-speaking participants. The fact that we find the same pattern independent of participants' language (or nationality) speaks for the robustness of our findings across Western countries. Yet, replications in different cultures are recommended as cultural differences in the appraisal of such pictures can be expected (cf., Kurdi et al., 2017).

Last, we focused exclusively on valence ratings in this research. In general, valence is considered to be the most important dimension in affective experiences (Lang & Bradley, 2007). However, classic theories on personality would also predict systematic interindividual differences in arousal ratings (H. J. Eysenck & Eysenck, 1985). Because valence

and arousal are not independent dimensions (i.e., arousal is higher for pictures of positive and negative than of neutral valence)[4], a thorough investigation of arousal effects might also further our understanding of interindividual differences in valence ratings. Future research should therefore investigate interindividual differences in arousal ratings as well. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

Conclusion

In the present research, we investigated interindividual differences in picture evaluations from a standardized database. We show that all Big Five traits except Openness are associated with interindividual differences in valence ratings of positive, neutral, and negative pictures. These findings have important implications for research designs in psychology and point to possible problems for interpreting their results. At the same time, they demonstrate the role of the Big Five in interindividual differences in emotional experiences.

---

4) An anonymous reviewer pointed out that picture sets with equal arousal might have led to a set of neutral pictures slightly more arousing than typical for neutral valence. High-arousing neutral pictures are often very ambivalent (Schneider et al., 2016) and thus might not represent "real" neutral valence. To control for this, we added the z-standardized arousal ratings from the OASIS as a further predictor into our model. However, this model only revealed a small positive main effect of arousal, suggesting that our findings should be robust even when using less arousing pictures. Detailed results of this analysis are provided in the Supplementary Materials (Table D1).

Supplementary Materials

For this article the following Supplementary Materials are available via the Open Science Framework (OSF) repository (for access see Index of Supplementary Materials below).

- Pre-registration
- Raw data for the study and main analysis
- Raw data for the OASIS ratings of the 90 pictures
- Codebook with brief explanation of variables in rawdata
- List of pictures, instructions, screenshots
- All analyses within the paper and supplement
- Output from all analyses within the paper and supplement
- All Analyses without exclusions
- All Analyses without exclusions Output
- Tables and short explanations on additional analyses

Index of Supplementary Materials

Ingendahl, M., & Vogel, T. (2021). *Supplementary materials to "Stimulus evaluation in the eye of the beholder: Big five personality traits explain variance in normed picture sets"* [Pre-registration]. OSF. https://doi.org/10.17605/OSF.IO/92WSU

Ingendahl, M., & Vogel, T. (2022). *Supplementary materials to "Stimulus evaluation in the eye of the beholder: Big five personality traits explain variance in normed picture sets"* [Data, Codebook]. PsychArchives. https://doi.org/10.23668/psycharchives.6669

Ingendahl, M., & Vogel, T. (2022). *Supplementary materials to "Stimulus evaluation in the eye of the beholder: Big five personality traits explain variance in normed picture sets"* [Additional materials]. PsychArchives. https://doi.org/10.23668/psycharchives.6670

References

Aluja, A., Rossier, J., Blanch, Á., Blanco, E., Martí-Guiu, M., & Balada, F. (2015). Personality effects and sex differences on the International Affective Picture System (IAPS): A Spanish and Swiss study. *Personality and Individual Differences, 77*, 143–148. https://doi.org/10.1016/j.paid.2014.12.058

Augustine, A. A., & Larsen, R. J. (2015). Personality, affect, and affect regulation. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, Volume 4: Personality processes and individual differences* (pp. 147–165). American Psychological Association. https://doi.org/10.1037/14343-007

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., & Fox, J. (2019). *Package "lme4."* [R Package]. https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bresin, K., & Robinson, M. D. (2015). You are what you see and choose: Agreeableness and situation selection. *Journal of Personality, 83*(4), 452–463. https://doi.org/10.1111/jopy.12121

Canli, T., Zhao, Z., Desmond, J. E., Kang, E., Gross, J., & Gabrieli, J. D. E. (2001). An fMRI study of personality influences on brain reactivity to emotional stimuli. *Behavioral Neuroscience, 115*(1), 33–42. https://doi.org/10.1037/0735-7044.115.1.33

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology, 38*(4), 668–678. https://doi.org/10.1037/0022-3514.38.4.668

Czerwon, B., Lüttke, S., & Werheid, K. (2011). Age differences in valence judgments of emotional faces: The influence of personality traits and current mood. *Experimental Aging Research, 37*(5), 503–515. https://doi.org/10.1080/0361073X.2011.619468

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., & John, O. P. (2019). Das Big Five Inventar 2. *Diagnostica, 65*(3), 121–132. https://doi.org/10.1026/0012-1924/a000218

Decuyper, M., De Pauw, S., De Fruyt, F., De Bolle, M., & De Clercq, B. J. (2009). A meta-analysis of psychopathy-, antisocial PD- and FFM associations. *European Journal of Personality, 23*(7), 531–565. https://doi.org/10.1002/per.729

Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging, 23*(3), 558–566. https://doi.org/10.1037/a0012897

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.* Springer.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Finley, A. J., Crowell, A. L., Harmon-Jones, E., & Schmeichel, B. J. (2017). The influence of agreeableness and ego depletion on emotional responding. *Journal of Personality, 85*(5), 643–657. https://doi.org/10.1111/jopy.12267

Gray, J. A. (1981). A critique of Eysenck's theory of personality. In H. J. Eysenck (Ed.), *A model for personality* (pp. 246-276). Springer. https://doi.org/10.1007/978-3-642-67783-0_8

Gross, J. J., Sutton, S. K., & Ketelaar, T. (1998). Relations between affect and personality: Support for the affect-level and affective-reactivity views. *Personality and Social Psychology Bulletin, 24*(3), 279–288. https://doi.org/10.1177/0146167298243005

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods, 40*(2), 512–521. https://doi.org/10.3758/BRM.40.2.512

Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., Taylor, J. H., & Crites, S. L., Jr. (2013). On the automatic activation of attitudes: A quarter century of evaluative priming research. *Psychological Bulletin, 139*(5), 1062–1089. https://doi.org/10.1037/a0031309

Hoff, H., Beneventi, H., Galta, K., & Wik, G. (2009). Evidence of deviant emotional processing in psychopathy: A fMRI case study. *The International Journal of Neuroscience, 119*(6), 857–878. https://doi.org/10.1080/00207450701590992

PsychOpen GOLD

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*(3), 390–421. https://doi.org/10.1037/a0018916

Howell, R. T., & Rodzon, K. S. (2011). An exploration of personality–affect relations in daily life: Determining the support for the affect-level and affect-reactivity views. *Personality and Individual Differences, 51*(7), 797–801. https://doi.org/10.1016/j.paid.2011.06.020

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). Guilford Press.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., & Liddle, P. F. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry, 50*(9), 677–684. https://doi.org/10.1016/S0006-3223(01)01222-7

Kurdi, B., & Banaji, M. R. (2019). Attitude change via repeated evaluative pairings versus evaluative statements: Shared and unique features. *Journal of Personality and Social Psychology, 116*(5), 681–703. https://doi.org/10.1037/pspa0000151

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods, 49*(2), 457–470. https://doi.org/10.3758/s13428-016-0715-3

Lang, P. J., & Bradley, M. M. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (Vol. 29, pp. 70–73). Oxford University Press.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* University of Florida, Gainesville.

Larsen, R. J., & Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology, 61*(1), 132–140. https://doi.org/10.1037/0022-3514.61.1.132

Levine, S. M., Alahäivälä, A. L. I., Wechsler, T. F., Wackerle, A., Rupprecht, R., & Schwarzbach, J. V. (2020). Linking personality traits to individual differences in affective spaces. *Frontiers in Psychology, 11*, Article 448. https://doi.org/10.3389/fpsyg.2020.00448

Lommen, M. J. J., Engelhard, I. M., & van den Hout, M. A. (2010). Neuroticism and avoidance of ambiguous stimuli: Better safe than sorry? *Personality and Individual Differences, 49*(8), 1001–1006. https://doi.org/10.1016/j.paid.2010.08.012

Lucas, R. E., & Baird, B. M. (2004). Extraversion and emotional reactivity. *Journal of Personality and Social Psychology, 86*(3), 473–485. https://doi.org/10.1037/0022-3514.86.3.473

Maio, G. R., & Esses, V. M. (2001). The need for affect: Individual differences in the motivation to approach or avoid emotions. *Journal of Personality, 69*(4), 583–614. https://doi.org/10.1111/1467-6494.694156

Revelle, W., & Scherer, K. (2009). Personality and emotion. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (Vol. 1, pp. 304–306). Oxford University Press.

PsychOpen GOLD

Rusting, C. L., & Larsen, R. J. (1997). Extraversion, neuroticism, and susceptibility to positive and negative affect: A test of two theoretical models. *Personality and Individual Differences, 22*(5), 607–612. https://doi.org/10.1016/S0191-8869(96)00246-2

Schneider, I. K., Veenstra, L., van Harreveld, F., Schwarz, N., & Koole, S. L. (2016). Let's not be indifferent about neutrality: Neutral ratings in the International Affective Picture System (IAPS) mask mixed affective responses. *Emotion, 16*(4), 426–430. https://doi.org/10.1037/emo0000164

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Smillie, L. D., Cooper, A. J., Wilt, J., & Revelle, W. (2012). Do extraverts get more bang for the buck? Refining the affective-reactivity hypothesis of extraversion. *Journal of Personality and Social Psychology, 103*(2), 306–326. https://doi.org/10.1037/a0028372

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://doi.org/10.1037/pspp0000096

Stead, R., & Fekken, G. C. (2014). Agreeableness at the core of the Dark Triad of personality. *Individual Differences Research, 12*(4-A), 131–141.

Tamir, M. (2009). Differential preferences for happiness: Extraversion and trait-consistent emotion regulation. *Journal of Personality, 77*(2), 447–470. https://doi.org/10.1111/j.1467-6494.2008.00554.x

Tobin, R. M., Graziano, W. G., Vanman, E. J., & Tassinary, L. G. (2000). Personality, emotional experience, and efforts to control emotions. *Journal of Personality and Social Psychology, 79*(4), 656–669. https://doi.org/10.1037/0022-3514.79.4.656

Tok, S., Koyuncu, M., Dural, S., & Catikkas, F. (2010). Evaluation of International Affective Picture System (IAPS) ratings in an athlete population and its relations to personality. *Personality and Individual Differences, 49*(5), 461–466. https://doi.org/10.1016/j.paid.2010.04.020

Vogel, T., Hütter, M., & Gebauer, J. E. (2019). Is evaluative conditioning moderated by Big Five personality traits? *Social Psychological & Personality Science, 10*(1), 94–102. https://doi.org/10.1177/1948550617740193