Supplementary Material

Spread the Joy:

How High and Low Bias for Happy Facial Emotions Translate into Different Daily Life Affect Dynamics

Section 1 of the supplementary material provides detailed descriptions of the No Fun No Glory (NFNG) selection procedures and the facial emotion identification morph task. Section 2 presents the exact coefficients and significance levels of the main analyses, followed by the results of a simulation study that we used to assess the reliability of mlVAR for our specific sample size, number of time points and model specifications. In Sections 3-5 results are presented of the sensitivity analyses we performed to assess the robustness of our findings.

1. Procedures

Selection of participants for the ecological momentary assessments based on the screening survey

We based the selection of individuals with persistent anhedonia on a general pleasure item from the Domains of Pleasure Scale [1]. Inclusion criteria were: (1) low levels of pleasure compared to peers, that is, below the 25th percentile; (2) a loss of pleasure, that is, the pleasure level they had been experiencing during the past two weeks was lower than the level they considered normal for themselves; (3) persistent anhedonia, that is, a loss of pleasure persisting for more than two months. Exclusion criteria were: inability to complete an electronic diary three times a day, use of psychotropic medication, professional therapeutic treatment for psychiatric problems and pregnancy. As a tandem skydive was one of the possible interventions offered in the NFNG Project, unwillingness to perform a tandem skydive, factors that prevented safe participation in a tandem skydive (epilepsy, cardiovascular problems, visual or hearing impairments, loose prostheses, a height of more than two meters, a weight of more than 95 kg, incapability to raise legs 90 degrees), and prior experience with skydiving, bungee jumping or base jumping were exclusion criteria as well. This screening process yielded a group of 148 anhedonic individuals, of whom 28 no longer met the inclusion criteria when contacted, 22 refused to participate and 29 could not be included for other reasons, resulting in a group of 69 participants. For each participant included in the anhedonia group, we matched a control participant on age, sex and educational level. Inclusion criteria for the control group were: (1) at least moderately high pleasure levels compared to peers, that is, above the 50th percentile; (2) no loss of pleasure, that is, the pleasure level they had been experiencing during the past two weeks had to be equal to or higher than the level they considered normal for themselves. Exclusion criteria for the control group were identical to those for the anhedonia group. Of the 114 individuals we sent an invitation to participate in the control group one participant no longer met the inclusion criteria, 21 refused to participate and 23 could not be included for other reasons. For a detailed description of the NFNG study, see Van Roekel and colleagues [2,3]. Please note that the interventions of the NFNG project started only after the first month of momentary assessments (observation month) and the second facial emotion identification task, therefore they did not interfere with the present study in any way.

Compensation participants

Participants received 10 euros for completing the screening survey and additionally participated in a lottery (prizes: 1 travel cheque, 4 iPad minis, and 15 fashion cheques). Participants who were included in the intervention study received a further compensation of 75 euros after completion of the observation month (of which the data were used in the present study), 125 euros after the first intervention month, 200 euros after the second intervention month, and 50 euros for each of the two follow-up measures. To receive these compensations, participants were required to complete at least 80% of the momentary assessments, as well as all monthly questionnaires and blood samples, which were part of the larger NFNG study. A tandem skydive was part of one of the interventions in the NFNG study (data were not used for the present study) and the opportunity to perform a tandem skydive free of charge could also be considered a compensation.

Because the tandem skydive as well as our financial compensation may have attracted a sample with altered reward sensitivity, we checked this by comparing participants who were unwilling to perform a tandem skydive to participants who indicated to be willing or perhaps willing to perform a tandem skydive in the total screening sample (N = 2,937) on reward responsiveness [4]. We also compared the reward responsiveness of the participants who were selected for the intervention study and agreed to participate to those who were not selected or did not agree to participate. Reward responsiveness did not significantly differ (i.e., $p > .05$) between participants who were unwilling (N = 376) and those who were willing or 'perhaps' willing (N = 2,561) to perform a skydive, and also not between participants who were selected for the intervention study and agreed to participate (N = 138) and those who were not selected or did not agree to participate (N = 2,799). Therefore, our data hold no evidence that the tandem skydive or the financial compensation participants received in the intervention study attracted a sample with altered reward sensitivity.

Facial emotion identification task

We used a morph task developed at Radboud University Nijmegen, the Netherlands [5,6]. Stimuli consisted of movie clips that lasted 10 seconds and contained 100 frames depicting the gradual change (i.e., 'morph') from a neutral facial expression to one of four full intensity emotional expressions: happiness, sadness, anger or fear (for examples, see the task description at https://osf.io/9edkh/). The movies had a resolution of 256 x 256 pixels, and were created with FaceMorpher (Luxand Inc., Alexandria, VA, USA) from high quality pictures of six different actors (50% females) from the Radboud Faces Database [7]. Pictures were cropped with an ovoid frame and converted to grey scale to avoid distracting external cues. Four movies were created of each actor, that is, one for each emotional expression. The original task contained 48 movie clips, that is, twelve per facial emotion, whereas we used a shortened version of 24 movie clips, that is, six for each emotion. A previous study in a large sample of young adults [6] indicated that the emotion identification patterns and reaction times for the shortened version are highly similar to the ones reported for the original 48-video-clip version of the morph task [5].

The morph task was programmed in Inquisit 4 (Millisecond, Seattle, USA). The task started with the instruction that participants were about to see movies of faces gradually changing from neutral to emotional expressions. Participants were asked to press the space bar as soon as they were able to identify the emotion. After pressing the spacebar the stimulus movie disappeared and participants indicated the emotion they identified by clicking on one of the four emotion labels. After clicking 'next' a fixation cross appeared in the middle of the screen for 500 ms, followed by a new stimulus. The order of the movie clips was randomized for each participant separately. Before the start of the actual task participants were shown a complete 10-s example movie, followed by two practice trials. After the practice trials the instructions were repeated, followed by the actual task consisting of 24 trials.

For each participant the mean reaction time (RT) of correctly identified trials was calculated per emotion, resulting in RT Happy, RT Sad, RT Angry and RT Fearful. RTs were calculated only if participants correctly identified at least four out of six movie clips of a specific emotion, otherwise they were considered unreliable. This resulted in 2 missing values for RT Sad, 1 for RT Angry and 2 for RT Fear at T0; and 1 missing value for RT Sad, 2 for RT Angry and 2 for RT Fear at T2. None of the remaining RT scores reached the maximum value of 10,000 ms, which indicates that the participants always pressed the spacebar before the movie clips stopped. As mentioned in our description of the selection procedure of the high and low happy bias groups, we calculated happy bias scores for each participant by taking the average of their RT scores on the other emotions (RT Sad, RT Angry and RT Fearful) and dividing it by RT Happy. Happy bias was used to select participants for the high and low happy bias group.

2. Exact coefficients and significance levels of mlVAR network analyses and results simulation study

Table S1

Standardized Coefficients for all Paths in the Network

	JOY		POS		INT		SAD		IRR		WOR		NEG	
	High bias	Low bias	High bias	Low bias	High bias	Low bias	High bias	Low bias	High bias	Low bias	High bias	Low bias	High bias	Low bias
JOY $_{t-1}$	**0.161***	**0.090***	**0.101***	0.030	**0.098**	0.046	-0.021	0.007	**-0.060**	0.041	-0.039	-0.017	**-0.060**	0.007
POS $_{t-1}$	**0.105***	0.018	**0.200***	**0.122***	**0.104***	0.034	-0.028	-0.023	-0.007	-0.016	**-0.076***	-0.017	0.025	-0.005
INT $_{t-1}$	**0.100**	0.034	0.037	**0.066**	**0.102***	**0.113***	0.014	-0.046	-0.011	-0.036	0.016	-0.041	-0.021	-0.025
SAD $_{t-1}$	0.051	0.001	-0.002	0.010	0.016	0.007	**0.070**	**0.079***	0.035	0.041	0.051	0.059	0.043	0.047
IRR $_{t-1}$	-0.020	-0.058	-0.014	-0.002	0.031	**-0.054**	-0.001	0.028	0.041	**0.111***	-0.045	-0.023	0.007	0.045
WOR $_{t-1}$	-0.032	**-0.069***	-0.043	-0.040	0.026	-0.053	**0.095***	**0.112***	0.022	0.023	**0.217***	**0.274***	0.033	0.021
NEG $_{t-1}$	-0.034	-0.025	0.026	-0.034	-0.038	0.010	**0.096***	0.037	0.035	**0.078***	**0.059**	0.006	**0.143***	**0.109***

JOY = feeling joyful; POS = pleasant experiences; INT = feeling interested in the things around me; SAD = feeling sad; IRR = feeling irritated; WOR = worrying; NEG = unpleasant experiences.

Bold coefficients were significant at $p < .05$. Bold coefficients marked with an * were significant at $p < .01$ and bold coefficients marked with an * and marked grey were significant at $p < 0.001$.

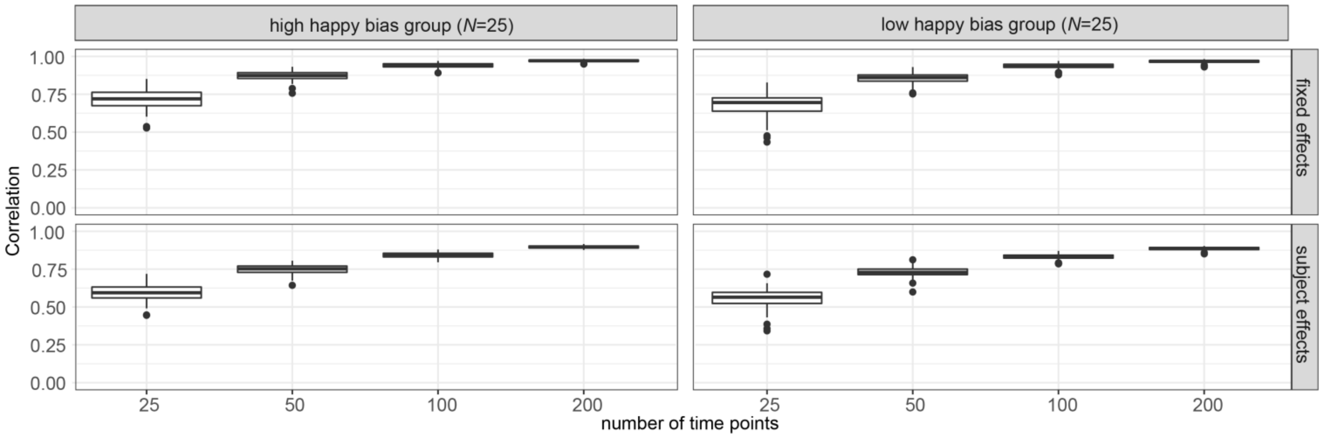

Figure S1. Results of a simulation study to assess the reliability of mlVAR temporal networks of seven nodes based on 25 subjects per group and a variable number of time points per person (25-200). In total 800 datasets were simulated, that is, 100 for each of the eight conditions (2 happy bias groups * 4 different time point conditions). The subject specific temporal networks found in the present study were generated as the true subject specific network structures. In order to assess how well the estimated networks resemble the true networks, we computed for each dataset the correlations between the true and the estimated subject specific temporal networks and between the true and the estimated fixed temporal networks. Boxplots indicate the distribution of the measures over all of the 100 simulated datasets per condition. In the top panel results are presented for the fixed effects and in the bottom panel for the subject specific effects (i.e., random effects), in the left panel for the high happy bias group and in the right panel for the low happy bias group.

3. Sensitivity analyses Mplus, multivariate models

A new version of Mplus (version 8) was launched recently. One of the new features is a package for dynamical structural equation modeling (DSEM) that uses a Bayesian estimator and allows the use of multivariate techniques for samples with a large number of assessments per participants and a small number of participants per group, like ours. We used this new package to perform a sensitivity check. As before, R packages qgraph version 1.4.4 and igraph version 1.1.2 were used to plot the networks and to compute and visualize the centrality indices.

Multivariate lag 1 network models were estimated for the low and high happy bias group separately by using the default specifications of DSEM, that is, Bayesian estimation with non-informative priors based on two independent Markov Chain Monte Carlo chains, and the potential scale reduction (PSR) criterion was used to assess convergence [8,9]. PSR < 1.1 was used as the default convergence criterion, but PSR < 1.05 has been recommended as well. For both bias groups the models converged after 800 iterations, that is, a PSR below 1.1 was reached. Recommendations to increase the number of iterations to check whether the PSR increased again were followed by using 20000 iterations. The PSR slightly increased and showed values > 1.1 after 800 iterations, but after, respectively, 2100 (high happy bias group) and 1800 (low happy bias group) iterations PSR < 1.05 was reached and the PSR remained stable and below 1.05 for the remainder of the 20000 iterations. This long sequence of low PSR values indicates model convergence. The network models and centrality indices reported in Figures S2 and S3 were based on 20000 iterations.

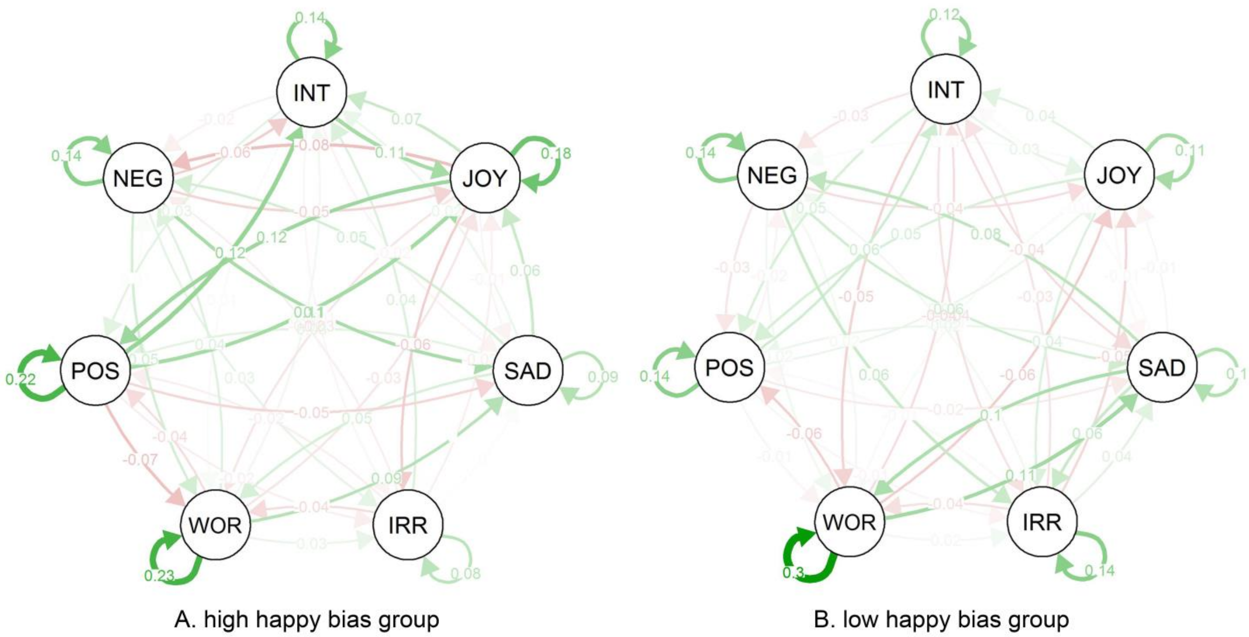

Figure S2. Complete networks high happy bias group (A) and low happy bias group (B), Mplus multivariate analyses. JOY = feeling joyful; POS = pleasant experiences; INT = feeling interested in things around me; SAD = feeling sad; IRR = feeling irritated; WOR = worrying; NEG = unpleasant experiences.

Figure S3. Centrality indices instrength and outstrength based on the complete network models, Mplus multivariate analyses. JOY = feeling joyful; POS = pleasant experiences; INT = feeling interested in things around me; SAD = feeling sad; IRR = feeling irritated; WOR = worrying; NEG = unpleasant experiences.

Even though this approach differed notably from the original one (e.g., a Bayesian estimator was used and a larger part of the information matrix was estimated by taking into account correlations between the dependent variables), the dynamical patterns are strikingly similar and the results confirm our main findings.

4. Sensitivity analyses for different group sizes

In the main analyses we used extreme happy bias groups of 25 participants in each group. Because there were no clear criteria on how extreme the groups should be, the exact number of individuals selected for each group was somewhat arbitrary. To assess the robustness of the results based on groups of 25 individuals we estimated the networks, computed the centrality indices, and performed permutation tests for bias groups of 20, 30, 35 and 40 individuals. We used the same mlVAR methods as in the main analyses.

For the groups of 25 individuals used in our main analyses permutation tests 1 and 2 reached statistical significance (see results section). We found that this was also the case for happy bias groups of *N*=20, *N*=30 and *N*=35 (see Table S2).

Table S2

Results Permutation Tests for Different Group Sizes

Group sizes of high and low happy bias group	Permutation test 1		Permutation test 2	
	Observed difference	*P*	Observed difference	*P*
N=20	0.52	0.015	0.28	0.011
N=25[a]	0.61	0.002	0.31	0.002
N=30	0.53	0.002	0.26	0.002
N=35	0.39	0.019	0.21	0.013
N=40	0.33	0.053	0.17	0.046

Note. [a] Main analyses; the others are sensitivity analyses.

Permutation test 1: Is the total (i.e., summed) absolute edge weight of all outgoing edges from JOY and POS, including the autoregressive edges, larger in the high happy bias group than in the low happy bias group?

Permutation test 2: Is the total edge weight of the outgoing edges from JOY and POS to JOY and POS, including the autoregressive edges, larger in the high happy bias group than in the low happy bias group?

For groups of 40 individuals the observed differences between the two happy bias groups were only half the size of those for the original groups of 25 individuals, and only permutation test 2 reached statistical significance. The results of these sensitivity analyses suggest that differences between the low and high happy bias groups become less pronounced as groups become less extreme. On OSF the network plots and instrength and outstrength plots are available for all groups (*N*=20 - *N*=40; https://osf.io/w823j/).

5. Anhedonia as a potential confounder

To assess the plausibility that anhedonia status may drive the results we found for happy bias we first tested whether the general affect dynamics we found to be associated with high versus low happy bias were also associated with anhedonia status. Permutation test 1, which was originally used to test whether the reward-related positive nodes JOY and POS more strongly predicted affect dynamics in the high happy bias group than in the low happy bias group, was repeated to explore whether a similar difference in affect dynamics could also be found between a control group and an anhedonia group. For the purpose of this sensitivity analysis, an anhedonia status variable was construed in which stability was taken into account. We started with the complete sample of participants who completed the first month of momentary assessments (N=138). Participants were assigned to the anhedonia group (N=60) only if their pleasure levels were low at T0 (< 25^{th} percentile) and remained low at T1 and T2, that is below the 35^{th} percentile, and they were assigned to the control group (N=59) only if their pleasure levels were high at T0 (> 50^{th} percentile) and remained above the 40^{th} percentile. Participants with less stability in anhedonia were treated as missing (N=19). We found that JOY and POS more strongly predicted affect dynamics in the control group than in the anhedonia group (observed difference = 0.286, p = 0.047). This suggests that at least partly similar affect dynamics are associated with happy bias and anhedonia status and that it is therefore important to adjust for anhedonia status.

To be able to adjust for anhedonia status, the subject-specific affect network paths (i.e., the random effects) based on the original VAR models were used to calculate the subject-specific total strength of the outgoing edges from JOY and POS. This resulted in a separate value for each participant of how strongly JOY and POS together predicted the other nodes in the network and themselves over time. Subsequently, we ran a linear regression analysis in SPSS version 25 with the total strength of all outgoing edges of JOY and POS per participant as outcome variable and dummy variables of anhedonia status and high versus low happy bias as predictors. Adjusting for anhedonia status still resulted in happy bias predicting the total strength of the outgoing edges from JOY and POS (β = .87, p < .001). This suggests that the effects we found for happy bias are not, at least not fully, driven by anhedonia status. For a plot of the individual variation in outstrength and instrength plotted separately for subjects with and without anhedonia, see Figure S4.

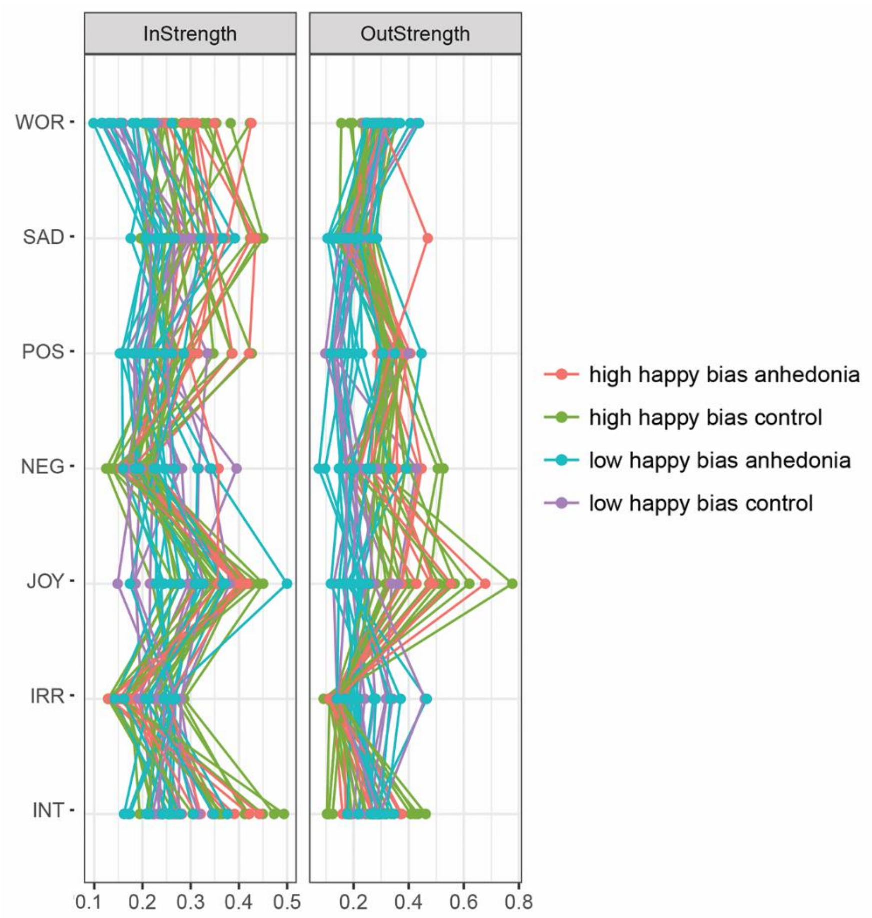

Figure S4. Individual variation in instrength and outstrength in four subgroups: anhedonic participants with a high happy bias (*n* = 8), anhedonic participants with a low happy bias (*n* = 13), control participants with a high happy bias (*n* = 17) and control participants with a low happy bias (*n* = 10). JOY = feeling joyful; POS = pleasant experiences; INT = feeling interested in things around me; SAD = feeling sad; IRR = feeling irritated; WOR = worrying; NEG = unpleasant experiences.

References

[1] M. Masselink, E. Van Roekel, V.E. Heininga, C. Vrijen, E. Nederhof, A.J. Oldehinkel, The Domains of Pleasure Scale (DOPS): A novel questionnaire to measure anhedonia, (n.d.) Manuscript submitted for publication.

[2] E. Van Roekel, M. Masselink, C. Vrijen, V.E. Heininga, T. Bak, E. Nederhof, A.J. Oldehinkel, Study protocol for a randomized controlled trial to explore the effects of personalized lifestyle advices and tandem skydives on pleasure in anhedonic young adults, BMC Psychiatry. 16 (2016) 182. doi:10.1186/s12888-016-0880-z.

[3] E. Van Roekel, C. Vrijen, V.E. Heininga, M. Masselink, E.H. Bos, A.J. Oldehinkel, An Exploratory Randomized Controlled Trial of Personalized Lifestyle Advice and Tandem Skydives as a Means to Reduce Anhedonia, Behav. Ther. 48 (2017) 76–96. doi:10.1016/j.beth.2016.09.009.

[4] I. Van den Berg, I.H.A. Franken, P. Muris, A New Scale for Measuring Reward Responsiveness, Front. Psychol. 1 (2010). doi:10.3389/fpsyg.2010.00239.

[5] G.M.A. Lodder, R.H.J. Scholte, L. Goossens, R.C.M.E. Engels, M. Verhagen, Loneliness and the social monitoring system: Emotion recognition and eye gaze in a real-life conversation, Br. J. Psychol. 107 (2015) 135–153. doi:10.1111/bjop.12131.

[6] C. Vrijen, C.A. Hartman, G.M.A. Lodder, M. Verhagen, P. de Jonge, A.J. Oldehinkel, Lower Sensitivity to Happy and Angry Facial Emotions in Young Adults with Psychiatric Problems, Front. Psychol. 7 (2016). doi:10.3389/fpsyg.2016.01797.
[7] O. Langner, R. Dotsch, G. Bijlstra, D.H.J. Wigboldus, S.T. Hawk, A. van Knippenberg, Presentation and validation of the Radboud Faces Database, Cogn. Emot. 24 (2010) 1377–1388. doi:10.1080/02699930903485076.
[8] A. Gelman, D.B. Rubin, Inference from Iterative Simulation Using Multiple Sequences, Stat. Sci. 7 (1992) 457–472. doi:10.1214/ss/1177011136.
[9] L.K. Muthén, B.O. Muthén, Mplus User's Guide, Muthén & Muthén, Los Angeles, CA, 1998.