**Supplementary materials for**

**Pleasant emotional feelings follow one of the most basic psychophysical laws (Weber’s law) as most sensations do**

**First &replication experiment**

**Experimental procedure**

The experiment’s code was written in Opensesame Web (Mathôt et al., 2012) and was run using JATOS (Lange et al., 2015) which is a web server that hosts online studies.

Participants enrolled to the studies using Sona (https://bgupsyc.sona-systems.com), which is a cloud-based research and participant management software. First, participants filled a depression questionnaire, which is a short, translated questionnaire taken from Psych Central website ([www.psychcentral.com](http://www.psychcentral.com)). We used an old version of the questionnaire that included 8 items that relates to the way participants felt and behaved during the past week. Only participants whose score fell within the non-pathologically depressed range (0-9) took part in the experiment. Additionally, for the first experiment, only participants who had a 13.9"-14.1" monitor in their laptop and a support for 1980X1020 resolution, could participate in the experiment, because we wanted to reduce the range of screen sizes as much as possible to reach near uniformity in circle sizes. We did not limit the monitor size in the replication experiment. In the pre-registration of the replication experiment, we declared that we would analyze the circle task only if at least 15 participants will have the same target circle size. Since less than 15 participants had the same target circle size we did *not* analyze the result of the circle task in the replication experiment. Additionally, we found in retrospect that the circle task is not a good reference point for the emotion task.

Both experiments began with detailed instructions. In addition, the experimenter provided a spoken explanation to ensure that the instructions are clear. Both experiments contained four practice blocks: a circle task practice block, emotion task practice block, key-mapping practice block, and an interleaved emotion-task – circle-task practice block (like in the experiment, proper), in that order. Participants were told that we are studying the influence of emotion on their circle perception; that we manipulate their emotion using pictures; and that we need to know how they personally feel in response to that picture. We emphasized that we are not interested in knowing how they are *supposed* to feel about the picture, or what most people would feel, but rather we want to know what the picture is *making them feel*.

After the instructions, participants opened the experiment online. Since, at the time, we did not know how to adjust the screen resolution based on monitor size using Opensesame Web, we needed to make sure that all information is seen on the participant’s screen. Most of the participants needed to use a zoom out function in order to make sure that they watch the full experiment-screen on their monitor. After that, participants were exposed to the target circle, and were asked to measure its diameter size. In the first experiment, we noticed that participants who had the same zoom-out level also measured an almost identical diameter size for the target circle. Since it is important for modeling purposes that all participants are exposed to the same circle sizes, we decided to conduct our analysis of the circle task only on the data of participants who used a 67% zoom-out level, which was the most common zoom-out value (*n=*17).

In both the circles task and the emotion task, we used a constant meaning-to-key mapping for all participants: "big" and "pleasant " were linked to a right key, and "small" and "unpleasant" were linked to a left key. The purpose of this was to make the meaning-to-key mapping as intuitive as possible, thus minimizing any possible response selection difficulty.

After the experiment was finished, the experimenter asked the participants three questions:

1. Please state in your words: what was your task in the experiment?
2. What strategy did you use in order to succeed in this task?
3. In your opinion, what was the goal of this experiment?

The purpose of asking those questions was to make sure the participant understood the instructions. After asking those questions, we sent the participants a video that was intended to induce an inspiring and positive mood (https://www.youtube.com/watch?v=7s22HX18wDY&feature=emb\_logo). The purpose of sending this video was to dilute the effect of the difficult-to-view images, so that the participants would leave the experiment in as-positive-as-possible mood. For the same reason, the experiment included more positive than negative emotional stimuli.

**Data exclusion**

We excluded data from the analyses based on an a-priori rules defined in the pre-registration. We conducted data exclusion separately for the circle task and the emotion task, using four steps for each task:

1. Relevant for the emotion task only: we excluded emotional stimuli whose more frequent responses did not fit the norms. Specifically, normatively pleasant stimuli (norm > 5) that were classified as pleasant by less than 60% of the participants and normatively unpleasant stimuli (norm<5) that were classified as unpleasant by less than 60% of the participants were removed from further analysis. In the first experiment, a total of 8 pictures were excluded from the analysis, and a total of 9 pictures were excluded from the analysis in the replication experiment.
2. We excluded participants from the analysis of a given task whose rate of normative responses was 2.5 SD less than the mean of the entire sample. In each one of the two experiments, one participant was excluded on this basis from the emotion task, and none was excluded from the circle task.
3. We excluded trials with RT greater than 7 sec, or shorter than 200 milliseconds.
4. We excluded trials with RT greater than 3.5 SD above mean RT (as determined per condition within participants). In the first experiment, the emotion task, a total of 34 trials were excluded, and for the circles task, 58 trials were excluded. In the replication experiment, 24 trials were excluded from the emotion task.
5. Relevant for the circle task only: only 17 participants were chosen for the later model analysis, based in their zoom-out level as describe above.

**Statistical analysis**

To assess comparative model fit, we used the Deviance Information Criterion (DIC, Spiegelhalter et al., 2014) as estimated per participant. DIC is a comparative index, with smaller values indicating a better model fit, while considering model’s complexity.

For our original analysis in the first experiment, we fitted 7 models for each task. The first model was the NULL model with all its parameters set to be equal across different levels of the independent variable in question (circle size for the circle task, and valence category for the emotion task). Then, we created five additional models, such that in each one of these models, a different parameter was set as a function of the independent variable (valence category). In the last model, both the drift rate (DR) and the drift rate standard deviation (SV) were set as a function of the independent variables.

In both tasks, in both experiments, we used one set of priors for the NULL model, and another set of priors for the rest of the models. Specifically, the mean priors for the NULL model of the for the circle task they were A=0.8, B=4, t0=.2, DR.true=2 and DR.false=.2, and SV=1 set the scaling.

For all models that we fit, we assessed the sampling adequacy of the posteriors using Potential Scale Reduction Factor (PSRF). An example for this successful convergence can be seen in fig.S1.

**The emotion task results- original analysis first experiment**

In the first preregistration, we planned an analysis in which we did not look for positive/negative separately (as done in the article itself) but instead, examined DR as a function of whether it represented the accumulator for the normative or aberrant response. In ‘ggdmc’ terms, the BANOVA according to the first pre-registration was conducted on DR.true and DR.false, as dependent variables. The details of this analysis do not contradict those of the analysis reported in the paper, but they are less readily interpretable. Details follow:

After data cleaning, we found for the emotion task that mean RT=1.391, and average aberrant response rate =0.112.

The second model (M-2) that we fitted was a model in which DR for the normatively correct accumulator was set as a function of valence category, while the rest of the model parameters were constrained. As predicted, we found that as the valence category was farther away from the neutral point (5), the mean DR was higher (*BF10*=3.326e +19, Fig.S2.a). In the third model (M-3), SV for the normatively correct accumulator was set as a function of valence category. In contrast to our original prediction, we found that as valence became more positive (and lass negative) SV increased (*BF10*=4.555e +8, Fig.S2.b). In M-4, the parameter A was set as a function of valence category. In M-5, the B parameter was set as a function of valence category, and in M-6, non- decision time was set as a function of valence category. For models M-4 through M-6, we did not have any prediction regarding parameter value, and these models were created only to ensure that whatever differences we find is specific to a given parameter rather than reflecting spillover from another parameter of the model which has been fixed in the given model. M-7 was our preferred model. In it both the DR and SV for the accumulator representing the normative response was set as a function of valence category. Again, as predicted, as the valence category was farther away from the neutral point (5), the DR increased (*BF10*=1.058e +19, Fig.S2.c). Regarding SV, we found the same result as for M-2: as valence became more positive (and lass negative) SV increased (*BF10*=8.018e +30, Fig.S2.d).

In order to show that M-7 is the superior model as compared to the rest of the models, we used individual Deviance Information Criterion (DIC) values to compare between models within each participant. Based on this comparison, we had the proportion of participants for whom a given model outperformed another given model. We then conducted Bayesian analysis of proportions to show that M-7 outperformed M-2 and M-3 (*p-outperforming=*0.727, *BF10*= 7.597; *p-outperforming=*0.970, *BF10*=1,122,397.00, respectively). However, while comparing M-7 to M-1, although in more participants M-7 outperformed M-1 (*p-outperforming*= 0.667), the proportions analysis indicated an undecided result (*BF10*= 1.884517).   
**The circle task results- original analysis, first experiment**

After data cleaning, we found that mean RT=0.887, and aberrant response rate =0.116.

For the circle task, we fitted 7 models (TVH-M1 to TVH-M7), that are based on the exacted same logic as for M-1 to M-7, but the independent variable in question was circle size and not valence category. As predicted, we found that as circle size was farther from the target circle size, DR increased (*BF10*=7.453e +19, Fig.S3.a). Regarding SV, we predicted that SV would increase with circle size, but we did not find such a clear effect (*BF10*=9705.327, Fig.S3.b). Rather, it seems that SV increased with circle size only for small circle sizes (1-4), but not for large circle sizes (5-8).

**The circle task results- new analysis, first experiment**

We fitted two models: M-B, a model for ‘bigger’ accumulator, and M-S, a model for the ‘smaller’ accumulator. We fitted both models using the same method described for M-P and M-N. Regarding the DR, we found similar results for those found for M-P and M-N: for M-B we found that as circle size increased DR increased as well (*BF10* =7.550e +63; Fig.S4.a). For M-S, we found that as circle size increased DR decreased (*BF10* =3.549e +61; Fig.S4.a). Regarding SV, we found for both M-B and M-S, that SV increased with circle size for circles that were smaller than the target circle, and then decrease for circles that were larger than the target circle (*BF10* =813.510; *BF10* =13.793, Respectively; Fig.S4.b). A possible post-hoc explanation for the difference in the SV results in M-B and M-S compared to M-P and M-N might be the different nature of the two tasks. The circle task, in contrast to the emotion task, included a comparison to a standard, which was not part of the emotion task.

**Linear trend**

Besides conducting Bayesian repeated measures ANOVA, post- hoc, we decided to test linear trends as well. For this analysis, we multiplied, for each participant, the vector of the posterior parameter estimates by a vector of weights representing a linear contrast (-5,-3,-1,1,3,5). The resultant scalar was then subjected to a single Bayesian two-sided t-test to show that the mean across participants is significantly different from zero.

We found linear trend in all analysis except for one (as describe in the article). Table S1 shows the results of the linear trend analysis.

**Rating experiment**

To rule out the possibility that the results of the first and the replication experiments are biased by the fact that we employed emotion bi-polar Polish norms (Marchewka et al., 2014), we conducted two rating experiments that were meant to rule out the possibility that the normative rating are biased by (a) coming from Polish culture, and (b) being bi-polar in nature. 68 participants from the same pool as those who performed the replication experiment were requested to rate the randomly ordered emotional evoking pictures. The experiment was conducted using Qualtrics (www.qualtrics.com). We analyzed the results using linear regression model. This analysis was not the original analyses we committed to in the pre- registration, but in retrospect this analysis is less time-consuming and it provides us with the ability to answer our research questions.

**Rating experiment a**

The purpose this experiment was to test whether the normative rating of the emotional evoking picture of NAPS are Polish-culturally dependent. 34 participants were exposed to the same emotional evoking pictures that were used in the first and the replication experiments and were asked to rate them on one valence scale ranged from 1 (very negative) to 9 (very positive), with 5 being neutral.

For each picture we calculate its mean valence. Then, we fitted a simple regression model (across pictures) with the dependent variable being the original NAPS rating, and the X-axis being the new mean rating of the given picture as collected in this experiment.

A significant regression equation was found (*BF*= 2.862523e+82), with an R2 of 0.9443. The original NAPS rating increased in 0.99736 for every unit of the new rating, making the relationship between the old and the new rating almost a perfect linear relationship. We also checked whether the data met the regression assumptions, which seems to be true (Fig. S5). Although the intercept was significantly different from zero (*BF*=*7.40695e+22*, a= 0.52733) implying that the Israeli participants rated the pictures slightly more positively than the Polish participants in(Marchewka et al., 2014), the slope (b=0.99736) was nearly unity indicating that the rank ordering of valence is not dependent on whether the norms were collected in Israel or in Poland.

**Rating experiment b**

The purpose of the rating experiment b was to rule out the possibly that the results are biased by the fact that the norms employed the bipolar approach to valence (Russell, 1980). Therefore, we conducted an experiment similar to rating experiment a, except for the fact that participants (N=34) were requested to rate each emotional evoking picture on two valence scale: one of pleasantness ranging from 1 (zero pleasantness) to 7 (very pleasant), with 4 being moderately pleasant, and a second unpleasantness valence scale ranging from 1 (zero unpleasantness) to 7 (very unpleasant), with 4 being moderately unpleasant. This new rating method seems to follow the unipolar approach (Watson & Tellegen, 1985), since we treated emotional intensity separately for positive and negative emotions.

For each picture, we calculate the normative (mean) valence twice: Once for unpleasant and once for pleasant. Then we created two simple regression models similar to the one in the rating experiment a. In the first regression model, x was mean pleasantness and in the second regression model, x was mean unpleasantness.

For the pleasantness model, a significant regression equation was found (*BF*=5.213725e+43), with an R2 of 0.7887. The original NAPS rating increased by 1.518 units for every additional unit of pleasantness. For the unpleasantness model, a significant regression equation was found as well (*BF*=1.612155e+52), with an R2 of 0.8402. The original NAPS rating decreased by 1.04495 for every additional unpleasantness unit. In both models, all the regression assumption were met (Fig.S6).

It seems that the norms collected in the rating experiment b are strongly correlated with the original NAPS rating. This finding renders the possibility of biased conclusions resulting from improper norms unlikely.

A picture containing shape

Description automatically generated**Fig. S1.**

Trace plots per participant for M-P, replication experiment. The X axis represents sample index (sequential position) and the Y axis represents posterior log- likelihood. Each color represents a different Markov chain. The model convergence is detectable since for every participant, all chains searched in the same parameter space, more-or-less.

**Fig. S2.**

Chart, line chart

Description automatically generated

**a.** DR as a function of valence category as obtained from M-2. As valence category was farther from the neutral point (5), DR increased. **b.** SV as a function of valence category as obtained from M-3. As valence became more positive, SV increased. **c.** DR as a function of valence category as obtained from M-7. As valence category was farther away from the neutral point (5), DR increased. **d.** SV as a function of valence category as obtained from M-7. As valence became more positive, SV increased.

**Fig. S3.**

Chart, line chart

Description automatically generated

**a.** DR as a function of circle size, as obtained from TVH-M7. As valence category was farther from the target circle size, DR increased. **b.** SV as a function of circle size as obtained from TVH-M7.

Chart, scatter chart

Description automatically generated**Fig. S4.**

**a.** DR as a function of circle size, for both M-B and M-S. For M-B, as circle size increased, DR increased as well. For M-S, as circle size increased, DR decreased. **b.** SV as a function of circle size, for both M-B and M-S. For both models, SV increased with circle size for circle that are smaller than the target circle, and then decrease for circles that are larger than the target circle.

**Fig. S5**Chart, scatter chart

Description automatically generated**.**

Regression model assumption test for rating experiment a. **a.** Testing the linearity assumptions: The dependent variable are the original NAPS valence norms and the on the X axis is the newly collected Israeli valence normative ratings. It seems from the scatterplot that the relationship the Polish and Israeli norms is linear. **b.** Testing the normality assumptions using normal Q-Q plot. It seems that the model does meet this assumption, since the standardized residual fell on the same axis as the theoretical quantiles do. **c.** Testing the homoscedasticity assumption. On the y axis are the model residuals, on the X axis are the fitted values. It seems that the residuals are evenly distributed around the fitted value, so the model seems to meet the homoscedasticity assumption as well.

**Fig. S6.** Graphical user interface, chart, application

Description automatically generated

Testing the linearity, normality and the homoscedasticity assumptionsfor the rating experiment b. **a,b,c f**or model 1 (pleasantness) and in **d,e,f** for model 2 (unpleasantness). It seems that all the regression-related assumptions have been met in both models.

**Table*. S1.***

|  |  |  |  |
| --- | --- | --- | --- |
| Experiment | Model | Parameter | BF |
| first | M-P | v | 2.992861e+16 |
| first | M-P | SV | 1.656758e+13 |
| first | M-N | v | 4.239821e+18 |
| first | M-N | SV | 1.656758e+13 |
| replication | M-P | v | 2.397399e+18 |
| replication | M-P | SV | 824114586838 |
| replication | M-N | v | 7.261738e+20 |
| replication | M-N | SV | 0.638341 |
| First& replication | M-N | SV | 154.0542 |

Linear trend analysis results.

**SM References**

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies” (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*(2), 596–610. https://doi.org/10.3758/s13428-013-0379-1

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society*, *76*(3), 485–493.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235. https://doi.org/10.1037/0033-2909.98.2.219