**Distribution Agreement**



Esther Jung March 26, 2024

Dissociating the Role of Valence on Motivated Behavior and Subjective Experience

by

Esther Jung

Philip Kragel, PhD
Adviser

Department of Neuroscience and Behavioral Biology

Philip Kragel, PhD

Adviser

Kristen Frenzel, PhD

Committee Member

Andrew Kazama, PhD

Committee Member

2024

Dissociating the Role of Valence on Motivated Behavior and Subjective Experience

By

Esther Jung

Philip Kragel, PhD

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Neuroscience and Behavioral Biology

2024

Abstract

Dissociating the Role of Valence on Motivated Behavior and Subjective Experience
By Esther Jung

Although emotion is highly subjective, researchers have attempted to model it objectively; one of the most prevalent models in emotion research today is the bipolar valence-arousal model. This model posits that combinations of valence, or the pleasantness or unpleasantness produced by stimuli, and arousal, or the degree of activation, are what dictate our emotions. Valence is also crucial to reinforcement learning; positively valenced stimuli can act as rewards and negatively valenced stimuli as punishments, driving our behavior and decision-making processes. Additionally, past research suggests that humans are vulnerable to a negativity bias, in which negative entities have a greater impact than positive entities. The present study aims to elucidate how valence drives behavior and impacts subjective experience, as well as to determine whether positive and negative stimuli are represented at different levels of granularity. To this end, 179 participants recruited from Amazon Mechanical Turk completed an assay of questionnaires, a naturalistic reward task, and a naturalistic threat task, providing measures of behavioral approach, avoidance, and emotional experience. Our results provided additional support for the bipolar valence-arousal model, confirming that combinations of valence and arousal contribute significantly to emotional experiences. However, self-report data did not provide evidence of a negativity bias, as subjective ratings for aversive, negative stimuli as compared to rewarding, positive stimuli. We found that participants adopted a strategy of altering their decision-making following errors in rewarding but not aversive contexts. This study furthers our current knowledge of the interactions between valence, decision-making, and subjective emotional experiences. Additionally, as our study results did not align with all predictions about the negativity bias, we may come closer to better understanding the controversial existence of the bias itself.

# Dissociating the Role of Valence on Motivated Behavior and Subjective Experience

By

Esther Jung

Philip Kragel, PhD

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Neuroscience and Behavioral Biology

2024

Acknowledgements

I would like to thank Dr. Philip Kragel for his constant support and guidance throughout my time in his laboratory and throughout this project. I would also like to thank the other members of the ECCO Lab for their valued time, knowledge, and feedback on this project.

**Table Of Contents**

## Introduction

Emotional experience is highly subjective, and for decades, researchers have attempted to develop measures to quantify it objectively. One prevalent model of emotion is the bipolar valence-arousal model (Russell 1980; Barrett & Russell 1999), which posits that hedonic valence and arousal are each continuous dimensions of emotion that lie at the core of all emotional experiences. In this model of affect, every event can be described as having some degree of valence and arousal. Valence, defined as the pleasantness or unpleasantness produced by stimuli encountered in the environment, often dictates our emotions and life experiences (Russell & Barrett, 1999; Kauschke et al., 2019). On the other hand, arousal is defined as the degree of stimulation or activation. Varying combinations of arousal and valence are proposed to account for diverse emotional states; for example, feelings of fear may result from the evaluation of negative valence and high arousal, whereas feelings of excitement may come from states of positive valence and high arousal (Russell & Barrett, 1999).

Valence is crucial to everyday life and plays a significant role in our emotional experiences, even contributing to our mental health. The dysfunction of brain systems related to valence is associated with various mental disorders, such as anxiety, depression, schizophrenia, and other mood disorders (Taylor et al., 2017; Bell et al., 1997). Investigating abnormalities in these systems, such as when the brain is exposed to a lack of positive stimuli or excessively negative stimuli, may open more doors to how the valence system influences the clinical and physiological symptoms of mental disorders. Valence systems may even be crucial to investigate when developing new treatments for these disorders. For example, hyperresponsivity to negative stimuli and a decrease in positive valence functioning are a neurobehavioral hallmark of posttraumatic stress disorder (PTSD) (Ben-Zion et al., 2022). Understanding how valence systems function in healthy individuals is essential for delving into what can go awry to cause dysfunction. With this study, we hope to shed more light on the effects of valence on human behavior and subjective experiences to build upon our understanding of valence, emotion, and mental health.

A primary function of valence is its ability to drive learning and decision-making through reinforcement, and pleasant and aversive experiences can act as rewards and punishments, respectively. Some researchers posit that emotional events are reinforcing; we will work for experiences that elicit positive emotions and try to escape or avoid experiences that elicit negative emotions (Rolls, 2000). Emotion's crucial link to memory and learning helps facilitate future behavior. Emotion modulates the encoding and retrieval of memories; emotionally salient stimuli often induce a "pop-out" effect that selectively enhances attention to those stimuli, and emotional states may be encoded along with memories, allowing for emotional states to trigger the recall of memories and thereby future decisions (Tyng et al., 2017; Rolls, 2000). As a result, decisions, such as approaching a reward or avoiding a threat, are influenced by a combination of valence and past individual experiences (Rangel et al., 2008).

Valence is linked to distinct neural mechanisms, from the level of the neuron to multi-modal circuits involving structures like the amygdala, nucleus accumbens, and insular cortex (Liu et al., 2011; Fullana et al., 2015). In particular, the orbitofrontal cortex (OFC) has been proposed to be a central brain region in sensory integration, emotional processing, and hedonic experience. The OFC integrates both sensory and motor information and directly connects to the basolateral amygdala, which likely contributes to goal-directed behavior. Activity in the medial

OFC is associated with the learning and memory of rewards, whereas lateral OFC activity is related to the evaluation of punishers (Kringelbach, 2005). Furthermore, the OFC is also linked to subjective conscious experience, which may contribute to its crucial role in evaluating the valence of stimuli (Kringelbach, 2005).

The main objective of the present work is to determine whether associations between motivated behavior and self-report vary as a function of hedonic valence. Past work has shown that humans are vulnerable to a negativity bias--the ability for negativity to have a greater impact on behavior than positivity, or the tendency to perceive negative entities (e.g., events, objects, personal traits) more strongly than positive ones (Rozin & Royzman, 2001; Norris, 2019). As a result, humans may feel a greater subjective emotional response to threatening stimuli than to rewarding stimuli, despite performing similarly across behavioral tasks with either type of stimuli. However, the existence of a negativity bias is highly contested; while some researchers have found behavioral evidence of a negativity bias and its neural correlates via neuroimaging, others have been unable to find support for the phenomenon (Norris, 2019).

Some scientists hypothesize that the negativity bias can be attributed to the differential processing of positive and negative information. The influence of negativity was tested in a 1991 study using a modified version of the Stroop task, in which words corresponding to positive and negative personality traits were displayed in various colors. When participants were asked to name the color of the word while disregarding the word itself, the researchers found that participants on average took longer to name the color of negative traits compared to positive traits, suggesting that the negative traits more strongly interfere with the color-naming task by taking up more attentional resources. In the same study, in a surprise free recall test, participants recalled about two times as many negative traits as positive traits (Pratto & John, 1991). Studies employing memory tasks have also revealed that participants' ability to discriminate between new items and old items (i.e., items they had already been exposed to) was better for negative words than positive words (Ortony et al., 1983; Robinson-Riegler & Winton, 1996). These results suggest that not only is negative information more attention grabbing, but negative stimuli also tend to be more easily remembered and recognized.

We hypothesize that positively and negatively valenced stimuli are processed distinctly by the brain, and that the negativity bias is supported by more robust engagement of brain systems to negative compared to positive stimuli. To evaluate this hypothesis, we aim to determine whether continuous dimensions of valence and arousal organize different types of emotional experiences. Previous research has shown that although a continuous dimension of valence does exist, it remains unclear to what extent arousal independently plays a role in emotion and how it relates to hedonic valence (Barrett & Russell, 1999).

Due to human tendency towards a negativity bias (Rozin & Royzman, 2001), we expect to see greater differences between self-report following negative, threatening stimuli and self-report following positive, rewarding stimuli (Norris, 2019). Under this account, participants will be more prone to feeling stronger emotions when seeing the negative stimuli compared to when seeing the positive stimuli. Further, negative emotions should be more distinct from one another than positive emotions. If this negativity bias is driven by an individual's reflective evaluation of their emotional state, rather than being driven by motivation or emotion, then behavioral data will show that participants' decision-making will be similar on both positive and negative valence tasks. As a result, there should be weaker correlations between self-report and avoidance behaviors, as compared to stronger correlations between self-report and approach behaviors. Alternatively, the negativity bias may reflect differences in motivational tendencies to approach and avoid stimuli, in which case experiential and behavioral measures will be highly correlated.

## Methods

### Participants and experimental procedure

Participants were recruited via Amazon Mechanical Turk, an established online recruitment platform for study participants. A total of 179 participants completed two behavioral tasks and three self-report questionnaires (demographic, personality, and neuropsychiatric questionnaires), including the Generalized Anxiety Disorder 7 (GAD-7) and Patient Health Questionnaire 9 (PHQ-9) (Table 1).

The two behavioral tasks were administered using PsychoPy and comprised a reward responsiveness and an acute threat task (Table 2). These tasks aim to measure participants' motivation towards rewarding positive stimuli and away from threatening negative stimuli, and the tasks are an extension of Delgado et al.'s reward processing task in which participants receive monetary rewards or penalties after choice-making (2000). The structure of the two tasks is the same: a card with an unknown value between 1 and 9 appears on the screen, and participants are asked to guess whether the value of the card is less than or greater than 5 by pressing the left or right keys. After a 2.5 second decision period, the value of the card is revealed for 0.5 seconds. Based on their response, participants view naturalistic video stimuli as feedback to their response (Figure 1), which were predetermined independently of participants' responses to evenly distribute the type of videos shown. During a win trial (when they guessed correctly) of the reward task, participants viewed a pleasant, rewarding stimulus, and during a loss trial (when they guessed incorrectly), they viewed a neutral stimulus. During the win trial of the threat task, participants viewed a neutral stimulus, and during the loss trial, they viewed an unpleasant, aversive stimulus. A neutral trial, when the value of the card was equal to 5 and the participant could neither win nor lose, resulted in the presentation of a neutral video in both the reward and threat task.

The naturalistic video stimuli in the two tasks were sourced from a repository of emotionally evocative short videos from Cowen and Keltner, 2017. For our study, we utilized 52 neutral videos, 70 rewarding videos, and 54 aversive videos sampled from the original repository of 2,185 videos. The video stimuli were assigned normative emotion categories based on participants' judgments of emotional states elicited by the videos (Cowen & Keltner, 2017). The video stimuli in our study belonged to one of 13 emotion categories: adoration, aesthetic appreciation, amusement, anxiety, boredom, calmness, craving, excitement, horror, interest, joy, nostalgia, and sexual desire. Videos categorized as adoration, craving, excitement, joy, or sexual desire were used as rewarding stimuli. Videos categorized as aesthetic appreciation, amusement, boredom, calmness, interest, or nostalgia were used as neutral stimuli. Videos categorized as anxiety or horror were used as aversive stimuli. After viewing the stimulus, participants are asked to report levels of "pleasantness", "anxiety", "activation", and "effort" on a visual analog scale ranging from 0 to 100 (Figure 2).

**Operationalization of conditions and behavioral variables**

We coded trial type (rewarding, neutral, or aversive) a nominal variable based on what kind of stimulus video the participants viewed. The pleasantness ratings were termed "positive" self-reported ratings, and the anxiety ratings were termed "negative" self-reported ratings. Participants' behavior was quantified with two variables. The first was a binary variable termed "switching," measured by differences in left or right key pressing when guessing the value of the card between one trial and the previous trial. If the participant stayed on the same key for two subsequent trials (e.g., pressed the left key two times in a row), switching was quantified as 0. If the participant switched keys between two subsequent trials (e.g., pressed the right key for one trial, then pressed the left key in the next trial), switching was quantified as 1.

To compare decision-making behavior across tasks, we created a new variable termed "motivationally consistent behavior," which was determined by whether switching behavior was consistent with the type of stimulus presented. We hypothesized that when participants were presented with an aversive stimulus, they would switch keys on the next trial to avoid another aversive stimulus, while when presented with a rewarding stimulus, participants would press the same key on the next trial in an attempt to obtain another rewarding stimulus. Accordingly, switching after an aversive trial and staying after a rewarding trial was termed "consistent" and quantified as 1; switching after a rewarding trial and staying after an aversive trial was termed "inconsistent" and quantified as -1. For all neutral trials, motivationally consistent behavior was quantified as 0, regardless of what key the participants pressed.

**Statistical Analysis**

We conducted statistical analyses on the behavioral choice data and self-report data using ANOVAs and linear models to determine the effect of trial type (neutral vs. negative and positive vs. neutral) and task (reward vs. threat) on self-report and behavioral variables. We also performed mixed effects models to determine whether associations between choice behavior and self-reported ratings varied depending on the task. To evaluate our hypothesis, we tested for distinct differences within self-report and behavioral variables when comparing the effect of trial type and task. All analyses were conducted in R (version 4.3.0).

**Distance-based analysis of self-report and choice data**

We fit regression models to assess the effect of valence on self-report and behavior and the similarity of self-report and approach/avoidance behavior using distance-based regression (Kriegeskorte et al., 2008; Kragel et al., 2018). To examine differences in the granularity of positive and negative emotions, we computed the Euclidean distance in average positive and negative ratings for each emotion category in each task. The observed distances were correlated with a model-based distance matrix based on the normative valence of stimuli (coding negative, neutral, and positive stimuli with values of -1, 0, and 1, respectively). Correlations between observed and model-based distance matrices were computed for each subject on each task. Inferences on correlation coefficients were performed using t-tests.

**Model verification and data visualization**

A principal component analysis (PCA) was used to align the self-report data into two components based on variance explained across experimental trials for each subject. Scores were plotted based on the average self-report ratings across trials for each subject and categorized by trial. Mean loadings were compared against zero using a t-test.

## Results

Descriptive statistics for self-report ratings for each task and trial type can be found in Table 3. Participants reported significantly higher positive ratings during the naturalistic reward task than the threat task, $t(3529.3) = 22.96$, $p < 2e\text{-}16$. Negative ratings were significantly higher in the threat task compared to the reward task, $t(3568.3) = 17.07$, $p < 2e\text{-}16$. Furthermore, when comparing the three types of video stimuli, participants reportedly felt less pleasant after viewing aversive stimuli than rewarding stimuli, $t(1356) = -22.83$, $p = 2.2e\text{-}16$. Participants also felt more negative after viewing aversive stimuli vs. rewarding stimuli, $t(1512.2) = 15.44$, $p = 2.2e\text{-}16$. On average, negative ratings after viewing aversive stimuli was not significantly higher than positive ratings after viewing rewarding stimuli, $t(1277.3) = -8.41$, $p < 2.2e\text{-}16$.

ANOVAs revealed that trial type significantly affected both positive (pleasantness) self-reported ratings (Table 4A) and negative (anxiety) self-reported ratings (Table 4B). Average positive and negative ratings also differed significantly across all three trial types, as indicated by paired t-tests comparing the positive and negative ratings within each trial type and between trial types (Figure 3).

Mixed effects models revealed that the main effect of task type was not significant on switching probability ($\beta = 0.11$, SE = 0.068, $p = 0.12$), but trial type had a significant effect on switching ($\beta = -0.25$, SE = 0.10, $p = 0.013$). The effect of negative ratings on switching behavior

($\beta = 0.00071$, SE = 0.0012, $p = 0.545$) was slightly weaker than the effect of positive ratings ($\beta = -0.0015$, SE = 0.0013, $p = 0.27$). On average, as positive ratings increased, switching behavior decreased, whereas switching behavior increased as negative ratings increased (Figure 4A).

The main effect of task type was significant on motivationally consistent behavior ($\beta = 0.22$, SE = 0.083, $p = 0.0081$), but the main effect of trial type was not as strong on motivationally consistent behavior ($\beta = 0.14$, SE = 0.10, $p = 0.18$). Similar to the effects on switching behavior, the effect of negative ratings on motivationally consistent behavior ($\beta = -0.00086$, SE = 0.0012, $p = 0.502$) was slightly weaker than the effect of positive ratings ($\beta = 0.0014$, SE = 0.0015, $p = 0.36$). On average, as positive self-reported ratings increased, so did motivationally consistent behavior; as negative ratings increased, motivationally consistent behavior decreased (Figure 4B). Logistic regression models evaluating the relationship between motivationally consistent behavior and trial type suggested that the log odds ratio of performing an inconsistent behavior vs. a consistent behavior would increase by 0.136 if moving from an aversive to a rewarding trial (SE = 0.101, $p = 0.18$).

PCA revealed that self-report data could be reduced to two dimensions: one component (Component 1) a bipolar representation of valence and the other (Component 2) the level of arousal (Figure 5). Furthermore, the mean loadings of both the positive and negative self-report variables were significantly different from zero ($p < 2.2e-16$). Positive self-report had a loading of 0.68 for Component 1 and 0.63 for Component 2. Negative self-report had a loading of -0.44 for Component 1 and 0.49 for Component 2.

The distance-based regression analyses took into account the average positive and negative ratings for each emotion category and conducted pairwise comparisons between categories to generate distance matrices (Figure 6). The correlation between the model-based matrix and observed matrix was slightly weaker when evaluating positive ratings (0.654) than when evaluating negative ratings (0.667).

**Discussion**

We observed that positive and negative self-reported ratings were strongly associated with trial type, confirming that the naturalistic video stimuli had a significant effect on participants' emotional states. Statistical comparisons suggested that on average, participants reportedly rated aversive stimuli as less pleasant and more negative compared to rewarding stimuli. These results were consistent with our expectations that aversive stimuli evoked fewer positive feelings and greater negative feelings, while rewarding stimuli evoked greater positive feelings and less negative feelings. Accordingly, the naturalistic reward task had greater positive ratings and less negative ratings than the naturalistic threat task, again confirming the tasks had their intended effects.

However, the magnitude of negative ratings after viewing aversive stimuli was not significantly higher than the magnitude of positive ratings after viewing rewarding stimuli, $t(1277.3) = -8.41$, $p < 2.2e\text{-}16$. If the negativity bias were at play, we would expect participants to report stronger emotion ratings after seeing the aversive stimuli compared to seeing the rewarding stimuli. However, our statistical analyses suggested the opposite; as a result, we are unable to conclude that the negativity bias is playing a significant role in participants' self-reported feelings after viewing aversive and rewarding stimuli.

We had predicted that the negativity bias would contribute to negative ratings corresponding to a greater effect on behavior. However, mixed effects models suggested that the effect of negative ratings on both switching behavior and motivationally consistent behavior was weaker than the effect of positive ratings. Additionally, motivationally consistent behavior tended to decrease on average as negative ratings increased, opposing our predictions about the negativity bias. If the negativity bias was present, participants may be more inclined to avoid aversive stimuli as they experience greater negative emotions. Accordingly, we had expected that motivationally consistent behavior would increase with self-reported negativity, but our model suggested the opposite--the odds of performing motivationally consistent behavior increased from aversive trials to rewarding trials, suggesting that rewarding trials are associated with greater motivationally consistent behaviors. Thus, we cannot conclude that the negativity bias is present in the interactions between self-reported ratings and behavior.

Our PCA analysis allowed us to identify a two-dimensional space of emotion consistent with the bipolar valence-arousal model; Figure 5 displays the clustering of aversive trials (in red) closer to the negative valence bound and rewarding trials (in green) closer to the positive valence bound. This model of emotion is consistent with previous emotion literature, providing additional support for the two-dimensional circumplex model first proposed by Russell (1980).

In a PCA, the loadings of a principal component are used to identify the cosine of the angle of rotation relative to the original axes and can range from -1 to 1. Loadings also identify the extent to which a variable contributes to the principal component. Whereas a positive loading indicates that a variable contributes to some degree to the principal component, a negative loading indicates that its absence contributes to the principal component. Furthermore, the larger a loading's relative magnitude (i.e., closer to 1 or -1), the more important its presence or absence is to the principal component (Harvey & Hanson, 2022). Positive self-report had a positive loading for Component 1 whereas negative self-report had a negative loading for Component 1. These loadings indicated that the combination of both the presence of positive ratings and the absence of negative ratings is significant for Component 1. Based on these loadings, we interpreted Component 1 as an axis of valence, with the presence of positive ratings and absence of negative ratings contributing to greater valence (pleasantness).

One general limitation of the study was the use of an online platform like Mechanical Turk to recruit participants. Because the task is administered virtually, a prominent challenge is determining whether participants are attending to the task and making truthful choices, such as in the self-report questionnaires and the key pressing to guess the value of the card. One study comparing responses received from Mechanical Turk participants to participants recruited on campus or through online forums did find a significant difference in the responses from Mechanical Turk (Bartneck et al., 2015). This difference might be attributed to the lack of the presence of an experimenter when participants complete an online task, which may result in less pressure on the participant to complete the task truthfully or finish it to the end at all. However, Mechanical Turk is valuable in that many participants can be recruited in a much shorter time frame than in person. In addition, the population of participants on Mechanical Turk is much more diverse than participants recruited on college campuses, who often share many more demographic characteristics.

Another limitation may be the ecological validity of our valence-arousal model. Although our model of emotion could be generalizable to everyday emotional experiences, the model may be limited due to the set of brief stimuli videos used in the study. The videos used in our behavioral tasks were about 5 seconds long on average. Although our results showed that the type of video significantly impacted the subsequent choices made during the task, the stimuli's effects may have been too temporary to be in play in decisions made multiple trials later. As a result, utilizing this repository of videos to build our model may not take into account the longer-term consolidation of emotional experiences in real life.

In conclusion, we were able to corroborate the valence-arousal model of emotion using participants' self-reported positive and negative ratings. On the other hand, we were unable to conclude that the presence of a negativity bias played a role in participants' behavior and subjective experiences. Our analyses suggested that there was not a significant difference in both subjective self-report and behavioral variables in response to negatively vs. positively valenced stimuli. Our study's sample was recruited online from Amazon Mechanical Turk, which may have affected the quality of data that we collected; in the future, we hope to support our current findings with an in-person sample of participants. In addition, our project at hand did not contain a neuroimaging component to confirm whether distinct systems were activated during the processing of negatively and positively valenced stimuli. Moving forward, we hope to integrate functional magnetic resonance imaging (fMRI) results with participants completing tasks as they are being scanned.

## Tables and Figures

### Table 1. Demographic and self-report assessments

| Assay | Description |
|---|---|
| Sociodemographic Questionnaire | Assess demographic information, education, and social status |
| *PHQ-9* | Self-report assessment for symptoms of depression |
| *GAD-7* | Self-report assessment for symptoms of anxiety |

**Table 2. Experimental Task Descriptions**

| | |
|---|---|
| Naturalistic reward task | Participants watch dynamic videos of natural rewards (e.g., depicting appetizing food, positive social interactions, etc.). A card with an unknown value between 1 and 9 appears on the screen, and participants are asked to guess whether the value of the card is less than or greater than 5 by pressing the left or right keys. Then, the value of the card is revealed for 0.5 seconds. Based on their response, participants view either a rewarding video if they win or a neutral video if they lose. |
| Naturalistic threat task | Participants watch dynamic videos of natural threats (e.g., attacking snakes, dangerous insects, extreme heights, etc.). The structure of this task mirrors that of the naturalistic threat task, and participants view either a neutral video if they win or an aversive video if they lose. |

**Figure 1. Naturalistic Reward and Threat Task**

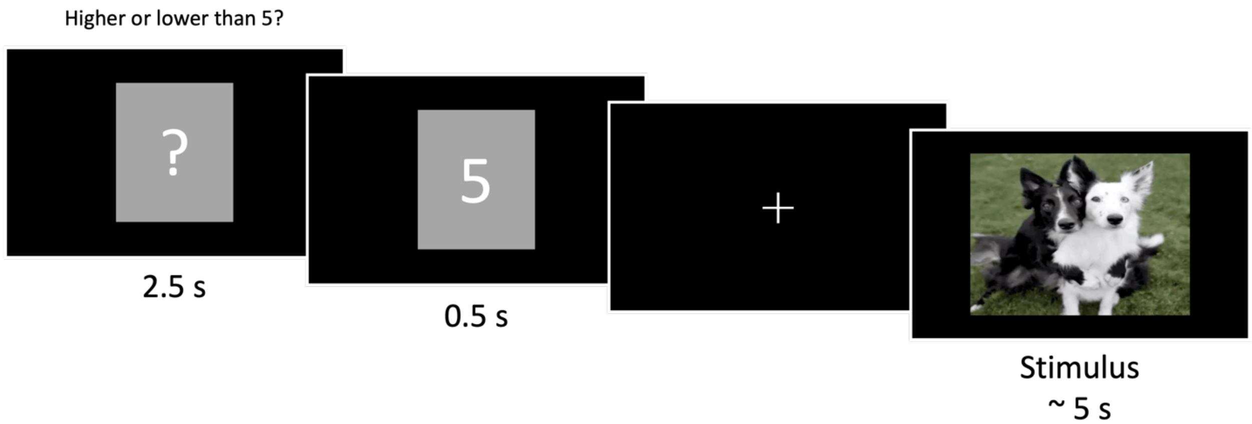



**Figure 2. Task Self-Report Questionnaires**

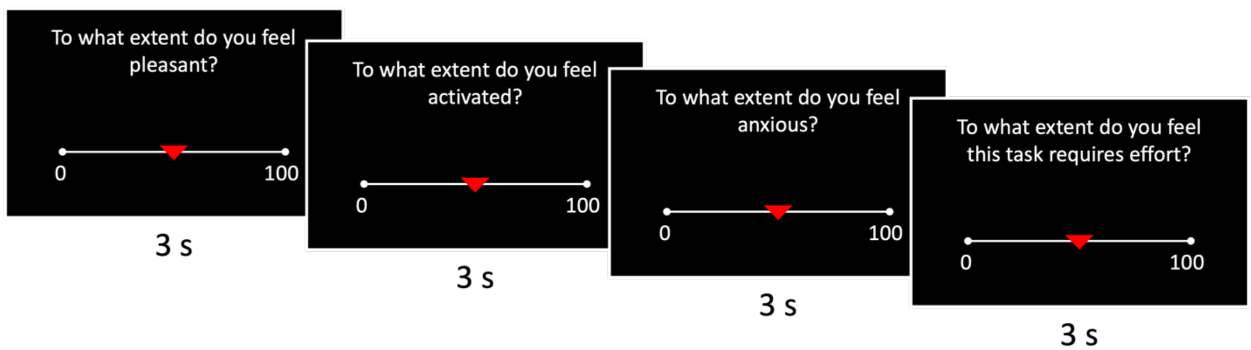


**Table 3. Descriptive statistics for positive and negative ratings by (A) task and (B) trial type.**

**(A)** Ratings by Task

| Task | | Positive ratings | Negative ratings |
|---|---|---|---|
| Naturalistic reward | Mean | 66.73 | 34.27 |
| | SD | 24.81 | 31.79 |
| Naturalistic threat | Mean | 47.53 | 51.26 |
| | SD | 27.40 | 30.84 |

**(B)** Ratings by Trial Type

| Trial Type | | Positive ratings | Negative ratings |
|---|---|---|---|
| Aversive | Mean | 39.82 | 56.00 |
| | SD | 28.25 | 30.91 |
| Neutral | Mean | 57.93 | 42.49 |
| | SD | 26.25 | 31.79 |
| Rewarding | Mean | 68.95 | 32.96 |
| | SD | 23.99 | 31.50 |

**Table 4. ANOVAs evaluating the effect of trial type on (A) positive ratings and (B) negative ratings.**

**(A)** Positive Ratings Affected by Trial Type

| | SS | df | $F$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 1130811.7 | 1 | 1658.308 | 0 |
| Trial Type | 343342.7 | 2 | 251.752 | 0 |
| Residuals | 2503279.6 | 3671 | | |

**(B)** Negative Ratings Affected by Trial Type

| | SS | df | $F$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 2235984.2 | 1 | 2246.621 | 0 |
| Trial Type | 213521.3 | 2 | 107.268 | 0 |
| Residuals | 3653620.0 | 3671 | | |

**Figure 3. Boxplot depicting average negative (red) and positive (blue) ratings across trial types and paired t-tests comparing positive and negative ratings for each trial type (*** p < 2.2e-16)**

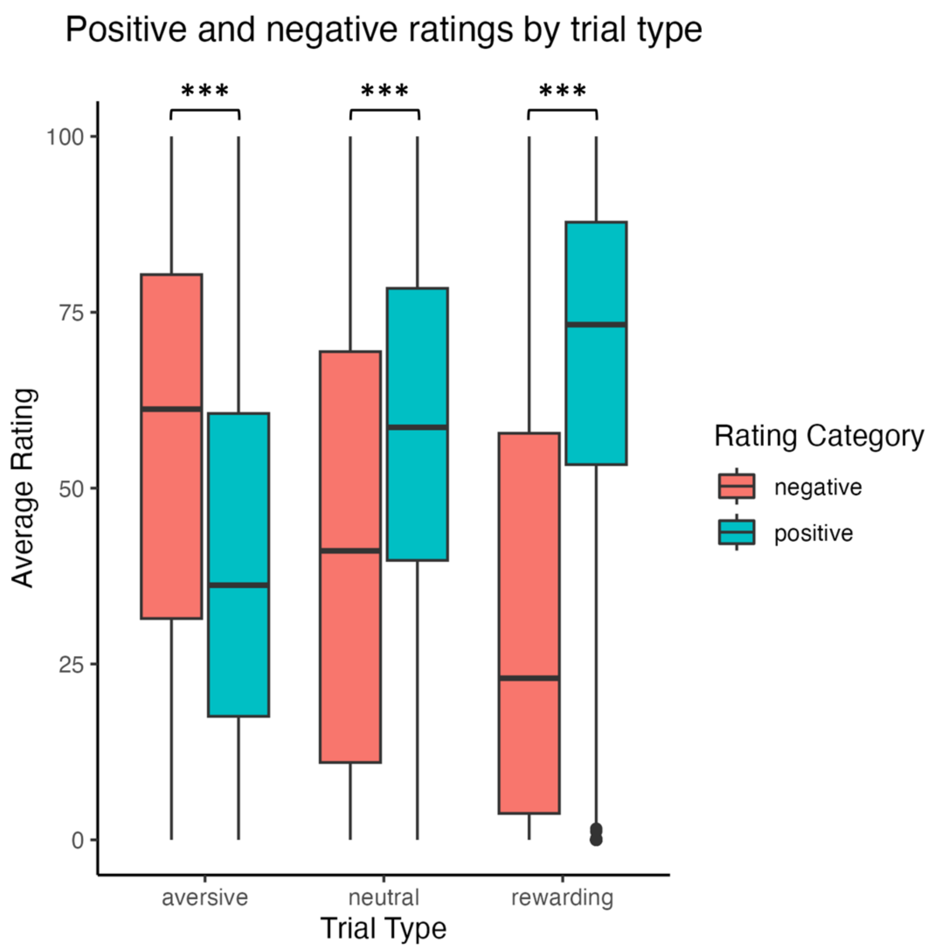


**Figure 4. Mixed effects models plot analyzing the effects of self-report ratings on (A) switching behavior and (B) motivationally consistent behavior. The shaded regions represent the 95% confidence intervals of the effects of the model at 0, 20, 50, 80, and 100 self-reported valence.**

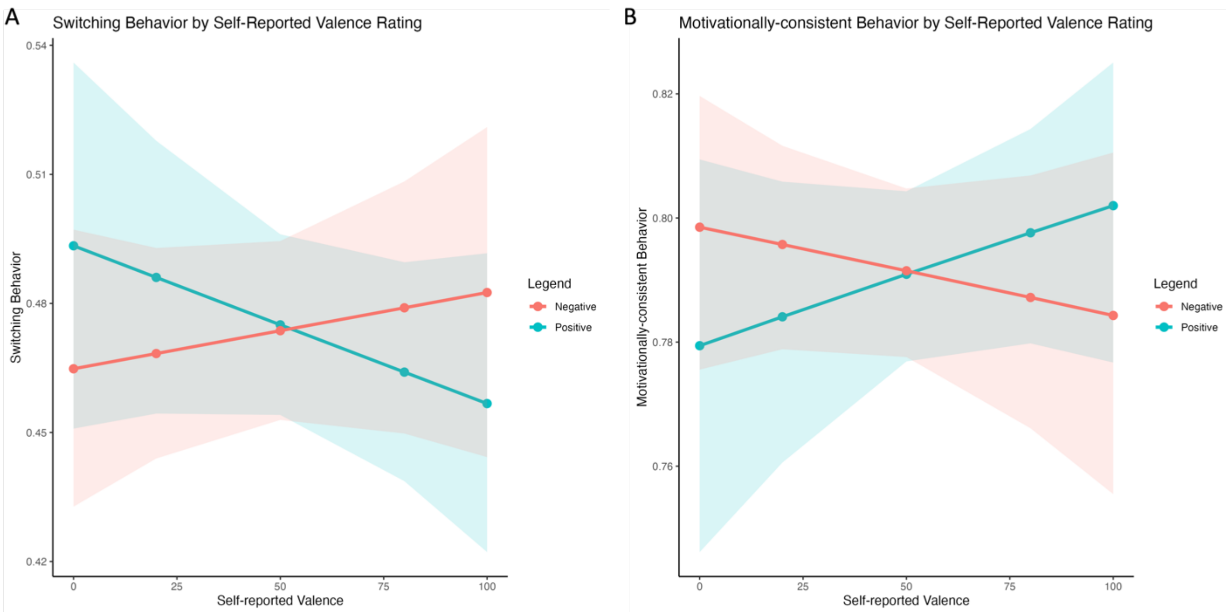


**Figure 5. Proposed two-dimensional model of core affect, plotted with scores of principal component analysis, based on self-report data and averaged by trial type**

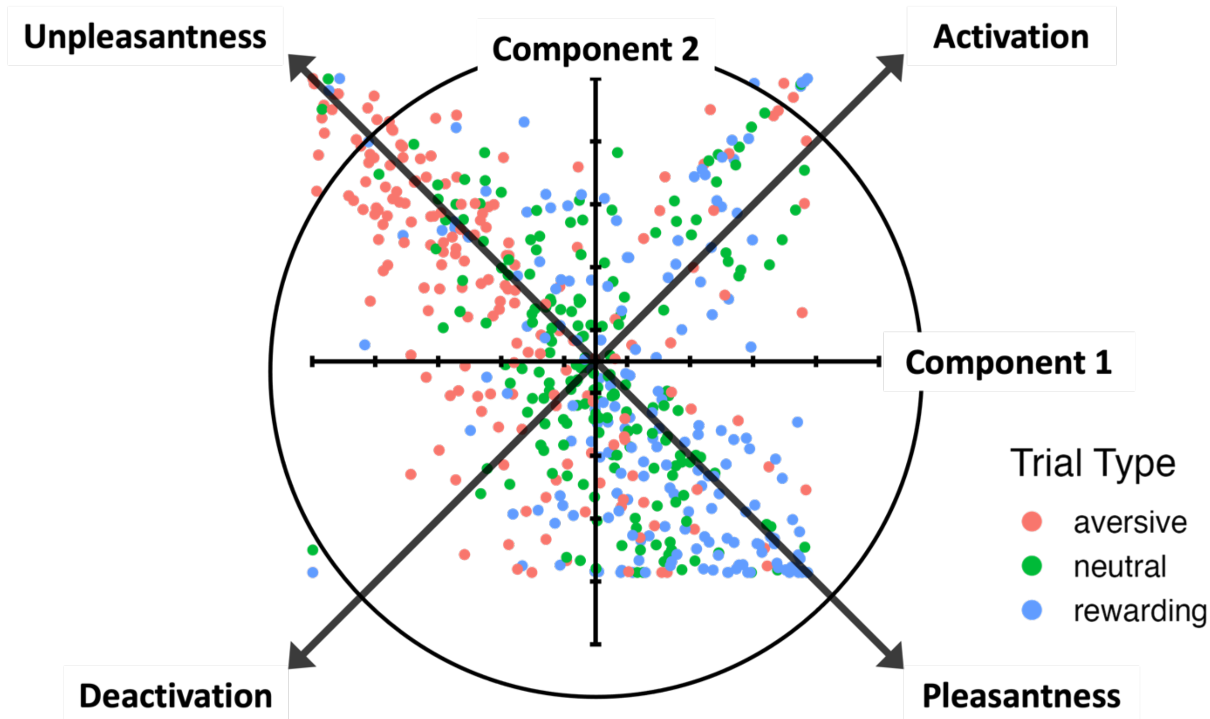
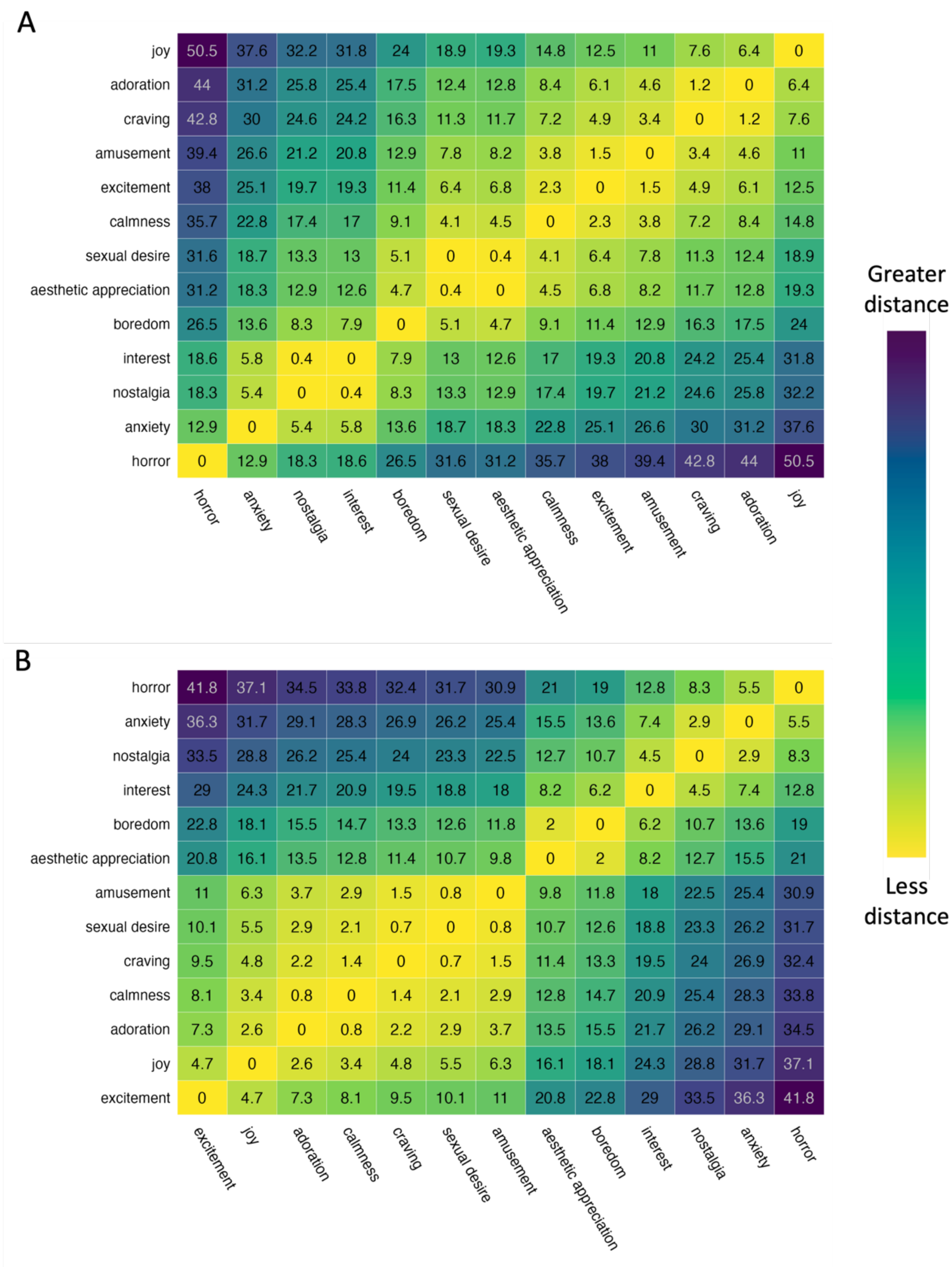


**Figure 6. Distance-based analysis of (A) positive self-report ratings and (B) negative self-report ratings. The 13 emotion categories were organized from lowest to highest average ratings for each category.**



**References**

Bartneck, C., Duenser, A., Moltchanova, E., & Zawieska, K. (2015). Comparing the similarity of responses received from studies in Amazon's mechanical turk to studies conducted online and with direct recruitment. *PLOS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0121595

Bell, M., Bryson, G., & Lysaker, P. (1997). Positive and negative affect recognition in schizophrenia: A comparison with substance abuse and normal control subjects. *Psychiatry Research*, *73*(1–2), 73–82. https://doi.org/10.1016/s0165-1781(97)00111-x

Ben-Zion, Z., Korem, N., Spiller, T. R., Duek, O., Keynan, J. N., Admon, R., Harpaz-Rotem, I., Liberzon, I., Shalev, A. Y., & Hendler, T. (2022). Longitudinal volumetric evaluation of hippocampus and amygdala subregions in recent trauma survivors. *Molecular Psychiatry*, *28*(2), 657–667. https://doi.org/10.1038/s41380-022-01842-x

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38). https://doi.org/10.1073/pnas.1702247114

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology*, *84*(6), 3072–3077. https://doi.org/10.1152/jn.2000.84.6.3072

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2015). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fmri studies. *Molecular Psychiatry*, *21*(4), 500–508. https://doi.org/10.1038/mp.2015.88

Harvey, D., & Hanson, B. (2022). *A conceptual introduction to PCA*. The Comprehensive R Archive Network. https://cran.r-project.org/web/packages/LearnPCA/vignettes/Vig_02_Conceptual_Intro_PCA.pdf

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–1236. https://doi.org/10.1016/j.neubiorev.2010.12.012

Kauschke, C., Bahn, D., Vesker, M., & Schwarzer, G. (2019). The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias? *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01654

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of Systems Neuroscience. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008

Kringelbach, M. L. (2005). The human orbitofrontal cortex: Linking reward to hedonic experience. *Nature Reviews Neuroscience*, *6*(9), 691–702. https://doi.org/10.1038/nrn1747

Norris, C. J. (2021). The negativity bias, revisited: Evidence from neuroscience measures and an individual differences approach. *Social Neuroscience*, *16*(1), 68–82. https://doi.org/10.1080/17470919.2019.1696225

Ortony, A., Turner, T. J., & Antos, S. J. (1983). A puzzle about affect and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 725–729. https://doi.org/10.1037/0278-7393.9.4.725

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*(3), 380–391. https://doi.org/10.1037//0022-3514.61.3.380

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Robinson-Riegler, G. L., & Winton, W. M. (1996). The role of conscious recollection in recognition of affective material: Evidence for positive-negative asymmetry. *Journal of General Psychology*, *123*(2), 93–104. https://doi.org/10.1080/00221309.1996.9921263

Rolls, E. T. (2000). On the brain and emotion. *Behavioral and Brain Sciences*, *23*(2), 219–228. https://doi.org/10.1017/s0140525x00512424

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

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037//0022-3514.76.5.805

Taylor, C. T., Lyubomirsky, S., & Stein, M. B. (2017). Upregulating the positive affect system in anxiety and depression: Outcomes of a positive activity intervention. *Depression and Anxiety*, *34*(3), 267–280. https://doi.org/10.1002/da.22593

Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01454

Unkelbach, C., Alves, H., & Koch, A. (2020). Negativity bias, positivity bias, and valence asymmetries: Explaining the differential processing of positive and negative information. *Advances in Experimental Social Psychology*, 115–187. https://doi.org/10.1016/bs.aesp.2020.04.005