Title:

Physiological Harmony or Discord? Unveiling the Correspondence Between Subjective Arousal, Valence and Physiological Responses

Authors: Alina Koppold[1], Tina B. Lonsdorf[1,2], Manuel Kuhn[1,3], Mathias Weymar[4], Carlos Ventura-Bort[4]

Keywords: Affect, Skin Conductance Response, Startle Blink Response, Valence, Arousal, Representational Similarity Analysis, Intraindividual Variability.

[1] Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2] Biological Psychology and Cognitive Neuroscience, Department of Psychology, University of Bielefeld, Bielefeld, Germany

[3] Department of Psychiatry, Center for Depression, Anxiety and Stress Research, Harvard Medical School, McLean Hospital, Belmont, Massachusetts, USA

[4] Department of Biological Psychology and Affective Science, University of Potsdam, Potsdam, Germany

Funding: This work was supported by an individual grant awarded by the German Research Foundation to TBL (DFG LO1980/4–1) as well as by grants awarded by the German Research Foundation in the context of the Transregional Collaborative Research Center CRC 58 with projects Z02 (INST 211/438–4) and B07 (INST 211/633–2 and INST 211/633–1) to TBL. This research was also supported by a grant from the Federal Ministry for Economic Affairs and Climate Action (EXIST program) to MW (03EGSSI013).

Credit Statement: AK and CVB were involved in the conceptualization, formal analysis, data curation, writing -original draft, and visualization, AK, CVB and MK were involved in investigation and data collection, MW and TBL were involved in funding acquisition and supervision. All authors participated in the writing-review and editing.

Abstract

Although affective state may occur almost inevitably together with physiological changes, it is still unclear whether events evoking similar affective experiences also produce comparable physiological responses or whether variation is the norm within individuals. To unravel this debate, we investigated the correspondence between physiological reactions and experienced arousal and valence, using representational similarity analysis. In two independent (Ns = 491 and 64) samples and three affect-inducing tasks, representational similarity matrices (RSMs) of skin conductance (SCR) and startle responses, as readouts of arousal and valence, respectively, were compared to RSMs of valence and arousal ratings that align with these competing theories. Furthermore, to address the role of intraindividual variability, we also investigated its influence on the relationship between physiological responses and experienced affect. We observed that dismissing vs considering intraindividual variability led to differential relations with models of valence and arousal, but models assuming variation proofed to be more representative of individual physiological patterns. Importantly, we observed strong to decisive evidence for a correspondence between SCR and startle responses and models of arousal and valence that assume variation, especially between trials evoking higher amplitude responses. These results emphasize the importance of considering intraindividual variability to assess the affect-physiology relationship (and in psychophysiology research in general) and invite to reconsider the notion of physiological reactivity as a potential indicator of affective states, prompting a shift towards understanding the mechanisms through which physiological changes contribute to conscious affective experiences.

Attending an electrifying concert, or watching a horror movie are experiences that evoke, almost ineluctably, affective subjective feelings accompanied by prominent bodily sensations (e.g., sweating, heart pounding, blood pressure drop). On a dimensional level, affect can be characterized by valence, referring to its hedonic value (ranging from pleasant to unpleasant), and arousal, referring to degree of activation (ranging from feeling calm to active, excited[1–7]). While there is a general agreement on the characteristics of affect[1,6], it is still unclear whether physiological changes mirror the experienced subjective affect. In other words, within individuals, do similar affective experiences produce comparable physiological reactions? Here, we aimed at answering this fundamental question with potential clinical implications[8–12], by investigating the intraindividual correspondence between affective subjective experiences and physiological reactions in two different samples[13,14] for three well-known affect-inducing tasks[15].

Previous scientific attempts on untangling the relationship between affective subjective experiences and associated physiological reactions have led to somewhat opposing views. Some views state that experienced affect is the by-product of the activation of two motivational brain circuits, the appetitive and the aversive systems, that serve survival functions for the individual and descendants[2,16–20]. Which circuit is activated, and their extent of engagement, would determine the degree of experienced valence and arousal, respectively. Importantly, when these circuits become active (e.g., in the face of life-sustaining or life-threatening events), a cascade of autonomic changes is orchestrated, preparing the organism for executing an adaptive overt response[20,21]. Although the validity of subjective affect as an indicator of the extent of activation of these circuitries is not unanimous[22–24], physiological reactions may markedly covary with evaluative reports of the experienced affect, given their coinciding underlying origin[18,20,25]. This covariation should translate into a strong intraindividual

correspondence between experienced affect and physiological changes. That is, in healthy individuals events evoking similar affective experiences should produce similar physiological responses. Because this assumption aligns in part with the *fingerprint* hypothesis of discrete emotions[26,27], which states that felt emotions should have consistent and unique physiological responses due to the existence of specific neural circuits[28], we also label it the *fingerprint* hypothesis of affect.[1]

Both in animals and humans, different psychophysiological measures have been pointed out as readout candidates of the activation of these motivational systems -and as correlates of the experienced valence and arousal. Two of the most commonly used measures are the skin conductance (SCR[29]) and the startle eye blink response[4,30,31]. There is compelling evidence that high arousing stimuli generally evoke higher SCR than their neutral low arousing counterparts[29,32,33]. This SCR pattern has been observed for both pleasant and unpleasant materials alike (e.g., images), indicating that SCR is insensitive to the hedonic value (i.e. valence insensitive), but indicative of the arousal level evoked by the triggering event[29,34,35]. On the other hand, the amplitude of the startle response has been associated with the induced valence level. Specifically, startle responses are potentiated across events categorized as unpleasant but attenuated while processing pleasant stimuli[29,36], with high arousing unpleasant/pleasant events evoking more potentiation/attenuation than low arousing ones[32,37–39]. These startle effects have been reliably replicated using different stimulus types from the perceptual domain[29,34,40–42].

[1] We refer to affect as a basic conscious experiences that can be reduced to the dimensions of valence and arousal. On the other hand, we refer to emotions as complex conceptual categories that emerge out of more basic psychological operations, including but not limited to valence and arousal characteristics[16, 87] .

More recent views propose that although experienced affect and physiological changes are interconnected, they are not influenced by the same underlying mechanism(s)[16,43–45,45]. These perspectives assume that autonomic reactions are strongly linked to the physiological needs of the current situation or context. Aligning with the principles of predictive coding, they further suggest that the brain is continuously trying to anticipate the required energetic needs of the body and execute the necessary physiological adjustments to fulfill them[46,47]. In turn, physiological changes associated with an incoming event do not reflect an isolated response but are dependent on the interplay between the physiological demands of the current context and the predictiveness of the encountered event. Importantly, these physiological reactions may partly subserve subjective affect. More concretely, experienced affect in response to a particular event may arise as the interaction between previous experiences with similar events, the ongoing read-out of the physiological bottom-up signals, and Bayesian inferences that determine both the confidence of the predicted event[44, 66, 88, 75] and the contrast between such predictions and the actual occurring event (i.e., prediction error)[89]. Crucially, due to the context and predictive heterogeneity of events categorized with similar affective attributes, the physiological needs in each of these instances may vary in a situated fashion. It is thus expected that within individuals, physiological variation is the norm across events evoking similar affective experiences. In other words, events producing comparable affective experiences may evoke different physiological reactions. This view is in line with the *populations* hypothesis of discrete emotions[26,27] (see also[51,52]) which states that emotions, as abstract concepts formed by heterogeneous instances, are not related to specific patterns of physiological activity. Thus, we refer to it as the *populations* hypothesis of affect. Recent reports have provided some evidence for within (and between) individual variability of physiological reactivity among events evoking similar affective experiences, supporting this hypothesis[26]. In one sampling study in which the relationship between subjective affective experience and physiological patterns

during a 2-week period was assessed, Hoemann et al.[26] observed heterogeneity rather than specificity in the relation between experienced valence and arousal and the evoked physiological patterns. Similar results are observed in more controlled, lab environments, using well-established affect-inducing tasks[15] and commonly used physiological measures[14].

Altogether, although prior studies pointed towards a strong correspondence between physiological reactivity and experienced affect[32,53], supporting the *fingerprint* hypothesis of affect, emerging evidence suggests situated variability, in agreement with the *populations* hypothesis[26]. To better understand these somewhat opposing results, some methodological considerations need to be contemplated. In many studies investigating a relationship between affective experience and physiological reactivity, findings are based on predefined categorical subdivisions (e.g., high vs low arousing) across the dimensions of arousal and valence. This approach implies artificially transforming dimensional constructs into subcategories, which could overshadow the inter-trial variability within categories for any given participant[90]. This transformation becomes critical to understand the relationship between affective experience and physiological reactivity because it could lead to ecological fallacy[54], by assuming that a relationship that occurs between groups of trials holds for individual trials. For instance, observing that the mean SCR is higher for a subgroup of events that evoke a high arousing experience, does not necessarily mean that a high arousing single event would evoke increased SCR than a low arousing event. Additionally, transforming dimensional variables into categories also prevents examining the association between affective experience and physiological reactions as a continuum, which becomes relevant to translate these findings to new understandings of psychological and clinical phenomena as dimensions[55]. Moreover, most of the studies have primarily focused on confirming the previously observed patterns (e.g., trials evoking high arousing states produce on average higher SCRs than their low arousing

counterparts) rather than testing whether other models may explain the affect- physiology association more accurately[56].

The question thus arises as to whether these considerations (i.e., disregarding intraindividual variability, lack of model testing) may pinpoint the correspondence between physiological reactions and affective experience in favor of one hypothesis over the other one. A suitable tool to test the influence of these factors in the affect-physiology relationship is representational similarity analysis (RSA[57]). Instead of grouping trials in predefined categories, RSA allows for calculating the so-called representational (dis)similarity matrices (RSMs) which depict the relationship between trials on a dependent variable (e.g., SCR), dimensionally (e.g., dimensionally close distance: two low arousing trials or e.g., dimensionally far distance: a low and a high arousing trial). Crucial to assess the role of intraindividual variability, RSMs can be calculated (1) taking into consideration inter-trial relationships within individuals -by first performing RSMs for each individual and then averaging them-, or (2) dismissing it -by first averaging responses across individuals and then calculating the RSM. Furthermore, RSA permits the creation of different distance models to test specific alternative hypotheses and determine which of them aligns with the data more closely[50]. Given the methodological advantages of this method, here RSA was used to test for the similarity between physiological reactivity and subjective affect, also considering different distance models of subjective affective experience. Particularly, the Nearest Neighbors (NN) model was used to test for evidence in favor of the *fingerprint* hypothesis (similar affective experiences evoke similar physiological responses). The NN assumes that a response on a specific trial is similar to responses on dimensionally close trials independently of their position on a dimensional scale (e.g., low, middle or high)[58], implying that events evoking similar affective experiences produce comparable physiological responses. Additionally, models that predict variation within a dimension that align more with the *populations* hypothesis (similar affective

experiences may evoke different physiological responses) were further used. These so-called Anna Karenina (AK) related models, which were recently introduced to assume variation of individual differences in brain-behavior relationships[58], predict similarities between dimensionally close trials on one side of a dimension, but dissimilarities on the other side of the dimension. More specifically, the AK model represents higher similarity between trials located in the higher extreme of a dimension but higher dissimilarity between trials from the lower end of the dimension (e.g., similar physiological responding across high arousing/valence trials, but dissimilar physiological reactions across low arousing/valence trials), whereas the inverted AK model represents the opposite pattern.

In a first step, the role of within-individual variation on the affect-physiology correspondence was assessed by testing whether (1) considering vs (2) removing intraindividual variability would produce different RSMs. These RSMs were then compared with NN and AK-related models of subjective affective experience. The *fingerprint* hypothesis of affect would be supported, if the RSMs of the SCR and startle show strong correspondence with the NN models of arousal and valence, respectively, after controlling for models depicting variation (i.e., AK-related models). On the other hand, the *populations* hypothesis of affect would be favored if the patterns of SCR and startle show correspondence with arousal and valence models representing variation (e.g., AK-related models), after controlling for competing models.

If removing vs considering intraindividual variability leads to differential associations with the NN and AK-related models, the question arises as to which RSM (with or without intraindividual variability) is more representative of the individual data. That is, do individuals show more often one pattern than the other? The answer to this critical point would inform about which of the two RSMs better represents the affect-physiology correspondence across participants. Hence, in a subsequent step, we tested the correspondence between the two RSMs

(with or without individual variation) and the RSMs from each participant to elucidate which of the two patterns offers a more common representation of the participants' physiological responses.

Box Terminology

1. **The *fingerprints* hypothesis of affect** assumes a strong intraindividual correspondence between experienced affect and physiological changes. That is, in healthy individuals events evoking similar affective experiences are expected to produce similar physiological responses. The term *fingerprint hypothesis* was initially employed in the field of discrete emotions[26,27] to group theories that assume that felt emotions should have consistent and specific physiological responses. Here, we decided to use the same term to refer to theoretical proposals advocating for consistent physiological reactions among events closed in the dimensions of valence and arousal.

2. **The *populations* hypothesis of affect** states that physiological variation is the norm across events evoking similar affective experiences. That is, events producing comparable affective experiences will evoke different physiological reactions. The term *populations hypothesis* was first used to group theoretical proposals stating that emotions, as abstract concepts formed by heterogeneous instances, are not related to specific patterns of physiological activity. Due to similarities with proposals assuming variation on the relation between physiology and affective subjective experience, we decided to use the same term to refer to the latter.

3. **Representational similarity matrices (RSMs) based on subjective affective experience** are created based on the valence and arousal ratings that participants provided for each event during the picture viewing, sound listening and imagery tasks. The formulae to calculate the similarities between trials depended on the model of interest: Nearest Neighbour (NN) model or Ana Karenina- (AK) related models).

    a. ***NN Model*** assumes that a response on a specific trial is similar to responses on dimensionally close trials independently of their position on a dimensional scale (e.g., low, middle or high). We used the NN model to test for the *fingerprint* hypothesis of affect.
    b. ***AK-related models*** assume that similarities between dimensionally close trials on one side of a dimension, but dissimilarities on the other side of the dimension. More concretely, the AK model represents higher similarity between trials located in the higher extreme of a dimension but higher dissimilarity between trials from the lower end of the dimension (e.g., similar physiological responding across high arousing/valence trials, but dissimilar physiological reactions across low arousing/valence trials), whereas the inverted AK model represents the opposite pattern. We used the AK-related models to test for the *populations* hypothesis.

4. **RSM based on physiological responses** are created based on the skin conductance (SCR) and startle eye blink responses. RSMs on physiological responses were calculated in two different ways:

a. ***Averaged RSM dismissing intraindividual variability.*** To calculate these RSMs, single trials are sorted by the variables of interest (e.g., arousal) and averaged across participants. Thereafter, the RSM of the averaged trials across participants is calculated.

b. ***Averaged RSM considering intraindividual variability.*** To create these RSMs, single trials are first sorted by the variable of interest (e.g., arousal), and then the RSMs of the physiological variable (e.g., SCR) calculated for each individual. Thereafter, individual RSMs are averaged together. This procedure was also used the create the averaged RSMs of subjective affect.

5. **Averaged level analysis** consists of the comparison between two averaged RSMs (e.g., averaged RSM of arousal based on the NN model and averaged RSM of SCR dismissing intraindividual variability). Averaged RSMs are compared, using Spearman's correlation. Thereafter, this similarity index is compared to 10,000 random permutations (i.e., permutation test) to test for evidence for a positive association.

6. **Individual level analysis** is performed by first comparing two RSMs for each participant, using Spearman's correlation (e.g., comparison between RSM of arousal based on the NN model and RSM of SCR for a specific subject), and then testing whether the similarity indexes are overall larger than 0, using simple t-tests.

7. **Regression analysis** is performed to test the relationship between two RSMs (e.g., averaged RSM of SCR dismissing intraindividual variability and averaged RSM of arousal based on NN model) after controlling for a third one (e.g., averaged RSM of arousal based on invAK model). RSMs (individual or averaged) are regressed out (e.g., the averaged RSM of arousal based on invAK is regressed out of the averaged RSM of arousal based on NN model) and the residuals are used to test for the similarity with the third RSM (e.g., the averaged RSM of SCR dismissing intraindividual variability) either using permutation tests (averaged level analysis) or testing whether the overall similarity indexes are larger than zero (individual level analysis).

Results

Skin Conductance Response (SCR)

To test the role of intraindividual variation on the affect-physiology correspondence, two averaged RSMs of SCR were created, one accounting for intraindividual (i.e., inter-trial) variability and the other dismissing intraindividual variability. Thereafter both averaged RSMs were compared to the averaged RSMs of arousal created based on NN and AK-related models. The NN model assumes that two events dimensionally close in a given variable (e.g., subjective arousal) evoke similar responses (e.g., SCR amplitudes), in line with the *fingerprint* hypothesis of affect, whereas AK-related models assume variations along the arousal dimension, aligning more with the *populations* hypothesis of affect. In the discovery sample (N= 483), which underwent a passive picture viewing task (PPV), both averaged RSMs of SCR (i.e., with and without intraindividual variability) were related to each other, rho = .08, p = .03, BF_{10} = 2.63, however, they showed differential relations with NN and inverted AK models of arousal.

For the averaged RSM of SCR in which intraindividual variability was disregarded, decisive evidence was found for a positive association with the averaged RSM of arousal ratings based on the NN model, and anecdotal evidence for a lack of a relationship with the averaged RSM of arousal based on the inverted AK model (NN model: rho = .55, p <.001, BF_{10} > 100; inverted AK model: rho = -.06, p = .89, BF_{10} = 0.37). However, it should be noted that the NN and the inverted AK models make similar predictions in the lower end of the arousal dimension but opposite predictions in the upper end of the dimension. Indeed, both models are related to each other (rho = .25, p <.001, BF_{10} > 100). Hence, to test the correspondence with the NN and the inverted AK models after controlling for the shared similarities, we regressed one model from the other out and compared the correspondence between the model residuals and the averaged RSM of SCR. This led to decisive evidence for a correspondence between

the averaged RSM of SCR dismissing intraindividual variability and the averaged RSM of arousal based on the NN model (rho = .58, p <.001, BF_{10} > 100), and decisive evidence for a lack thereof with the averaged RSM of arousal based on inverted AK model (rho = -.18, p =.999, BF_{10} =0.009; Figure 1A-C). These results indicate that, when no intraindividual variability is taken into account, the averaged pattern of the SCR is better explained by the NN model.

For the averaged RSM of SCR considering intraindividual variability, we observed decisive evidence for correspondence with RSMs of both models (NN model: rho = .25, p <.001, BF_{10} > 100; inverted AK model: rho = .62, p <.001, BF_{10} >100), indicating that the averaged pattern of the SCR was similar to both the NN and the inverted AK models of arousal. Interestingly, when the similarity between models was regressed out, we observed decisive evidence for a correspondence between the averaged RSM of SCR and the averaged RSM based on inverted AK model (rho = .57, p <.001, BF_{10} >100), and decisive evidence for the lack thereof with the averaged RSM based on NN model (rho = -.02, p =.887, BF_{10} =0.003; Figure 1D-F). These results suggest that, when intraindividual variability was considered, the averaged pattern of the SCR was strongly related to the inverted AK model of arousal.

Given that both averaged RSMs of SCR (dismissing and considering intraindividual variability) showed differential associations with the NN and inverted AK models of arousal, in a follow-up analysis, we tested, which of the two averaged RSMs of SCR was a better representative of the individual RSMs of arousal. Results revealed evidence for a lack of a relationship between the individual RSMs of SCR and the averaged RSM of SCR dismissing intraindividual variability, Mean= .001, $t(483)$= 0.62, p = .53, BF_{10} = 0.53, however, decisive evidence was observed for a positive association between the individual RSMs and the RSM of SCR considering intraindividual variability, Mean = .054, $t(483)$ = 12.34, p <.001, BF_{10} >100. These findings, which were further corroborated when models were regressed from each

other out (RSM of SCR without intraindividual variability: Mean = -.01, *t*(483)= -2.12, p = .029, BF_{10} = 0.55; RSM of SCR with intraindividual variability: Mean = .055, *t*(483)= 12.34, p < .001, BF_{10} >100; Figure 1G), indicate that participants more often show a pattern of SCR similar to the averaged RSM of SCR that considers the individual fluctuations in physiological responses over trials.

Additionally, individual level analysis was performed to test the relationship between individual RSMs of SCR and the individual RSM of subjective arousal based on the NN and the inverted AK models. Results showed evidence for a positive relationship of both the NN and the inverted AK models of subjective arousal with the individual RSMs of SCR (NN model: Mean = .014, *t*(482) = 3.93, p < .001, BF_{10} = 98.25; inverted AK model: Mean = .048, t(482)= 8.505, p <.001, BF_{10} >100). However, regression analysis revealed decisive evidence for a positive correspondence with the individual RSM of subjective arousal based on the inverted AK model (Mean = .047, *t*(482) = 8.72, p <.001, BF_{10} >100) and substantial evidence for a negative correspondence with the individual RSMs based on NN model (Mean = -.01, t(482)=-2.37, p = .018, BF10= 9.13; Figure 1H). Subsequently, we tested whether the evidence in favor of a relationship between the individual RSMs of SCR and the RMSs of subjective arousal based on the inverted AK model remained after controlling for other models, including valence- and arousal-based categorical, and valence-based dimensional models. In all cases, decisive evidence for a positive association between the inverted AK model and the SCR pattern was found after controlling for other models: relationship between SCR and inverted AK model of arousal after controlling for the valence-based categorical model, Mean = .044, t(482)= 8.08, p <.001, BF_{10} >100, the arousal-based categorical model, Mean = .046, t(482)= 8.23, p <.001, BF_{10} >100, and the valence-based dimensional model, Mean = .04, t(482)= 7.8, p <.001, BF_{10} >100; Figure 1I).

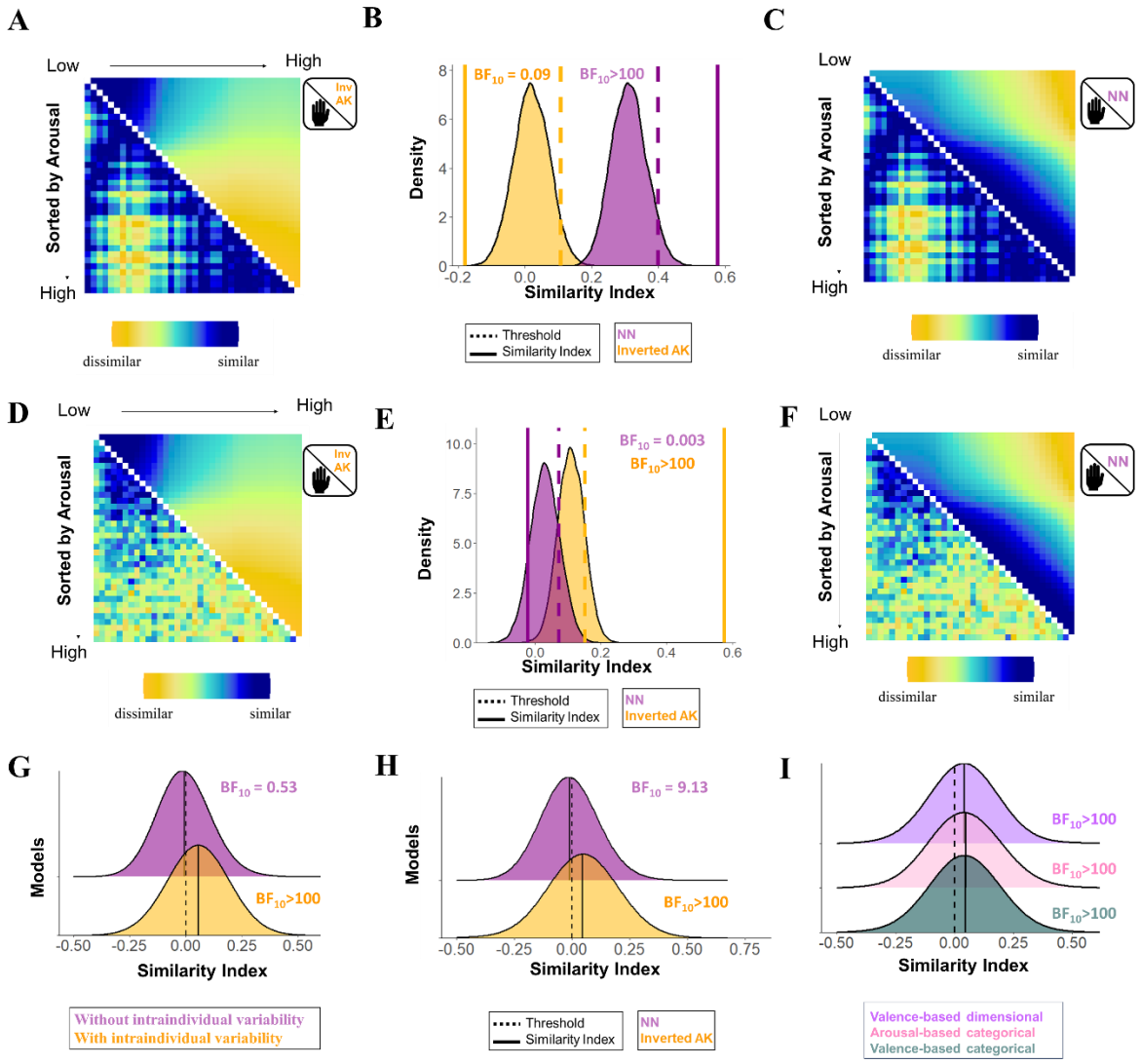

Figure 1. Association between the representational similarity matrices (RSMs) of subjective arousal based on the inverted Anna Karenina (inv AK) and Nearest Neighbours (NN) models (after controlling for the shared similarities), and the RSMs of the skin conductance response (SCR) dismissing (A-C) and considering (D-F) intraindividual variability (discovery sample, N= 483). A) Averaged RSM of SCR dismissing intraindividual variability (lower diagonal) and averaged RSM of subjective arousal based on the inverted AK model (upper diagonal). B) Results of the permutation test. C) Averaged RSM of SCR dismissing intraindividual variability (lower diagonal) and averaged RSM of subjective arousal based on the NN model (upper diagonal). D) Averaged RSM of SCR considering intraindividual variability (lower diagonal) and averaged RSM of subjective arousal based on the inverted AK model (upper diagonal). E) Results of the permutation test. F) Averaged RSM of SCR considering intraindividual variability (lower diagonal) and averaged RSM of subjective arousal based on NN model (upper diagonal). G) Distribution of the association between individual RSMs of SCR with the unique contribution of the averaged RSMs of SCR disregarding (purple) and considering (yellow) intraindividual variability. H) Distribution of the association between individual RSMs of SCR and unique contribution of the individual RSM of arousal based on the inverted AK and NN models. I) Distribution of the association between individual RSMs of SCR and individual RSMs of subjective arousal based on the inverted AK model after controlling for other models (i.e., valence-based categorical, arousal-based categorical, valence-based dimensional models).

The replication sample (N= 64), which underwent a PPV, passive sound listening (PSL) task and Imagery task, allowed us to test whether the significant association between the RSMs of subjective arousal based on the inverted AK model and the RSMs of SCR would also be reproduced in another sample and could be generalizable across tasks. Results revealed a similar pattern of results to the discovery sample when participants were instructed to passively view affective images (PPV), listen to affect-inducing sounds (PSL) and to imagine affectively-laden events (Imagery task; Suppl. S4-S6), indicating that the correspondence between SCR and the inverted AK model of arousal was independent of the specific task and replicable across samples.

Eye blink startle response.

The two averaged RSMs of startle eye blink (dismissing and considering intraindividual variability) were compared with averaged RSMs of subjective valence based on models that align with the *fingerprints* (i.e., NN model) and with the *populations* hypothesis (AK-related models) of affect. In the discovery sample, both averaged RSMs of startle eye blink (i.e., with and without intraindividual variability) were slightly related to each other rho = .21, p = .02, BF_{10} = 0.04, and showed disparate relations with NN and AK models of valence.

For the averaged RSM of startle blink responses in which intraindividual variability was disregarded, decisive evidence was found for a positive association with the averaged RSM of subjective valence based on the NN model, and anecdotal evidence for no relationship with the averaged RSM of valence based on the AK model (NN model: rho = .69, p <.001, BF_{10} > 100; AK model: rho = .05, p = .80, BF_{10} < .001). Subsequent regression analysis revealed decisive evidence for a correspondence between the averaged RSM of startle dismissing intraindividual variability and the averaged RSM of subjective valence based on the NN model (rho = .58, p <.001, BF_{10} > 100), and decisive evidence for a lack thereof with the averaged RSM based on

the AK model (rho = -.18, p =.999, BF_{10} =0.09; Figure 2A-C), indicating that, when no intraindividual variability was considered, the averaged pattern of the startle eye blink was better explained by the NN model.

When the averaged RSM of startle was computed considering intraindividual variability, an association was exclusively found with the averaged RSM of subjective valence based on the AK model (NN model: rho = .1, p = 0.32, BF_{10} = 2.03; AK model: rho = .62, p <.001, BF_{10} = 83.65; regression analysis: (NN model: rho =0, p =.91, BF_{10} = 0.004 ; AK model: rho = .63, p <.001, BF_{10} >100; Figure 2D-F).

Given the differential relationships between averaged RSMs of startle that dismiss vs consider intraindividual variability and the averaged RSM of subjective valence based on the NN and AK models we tested which of the two averaged RSM of startle was more closely related to the individual RSMs of startle. Results revealed no evidence for a positive relationship between the individual RSMs of startle and the averaged RSM of startle dismissing intraindividual variability, Mean = 0.007, *t*(398)= 2.5, p =.01, BF_{10} = 1.23, but decisive evidence for a positive relationship with the averaged RSM of startle that took into account intraindividual variability, Mean = 0.068, *t*(398)= 13.72, p <.001, BF_{10}>100 (regression analysis: RSM of startle dismissing intraindividual variability: Mean = -0.003, *t*(398)= -.98, p = .32, BF_{10} = 0.09; RSM of startle considering intraindividual variability: Mean = 0.068, t(398)= 13.86, p <.001, BF_{10} >100; Figure 2G), indicating that participants more often show a startle pattern similar to the averaged RSM that takes into consideration intraindividual variability.

Individual level analysis further demonstrated no association between individual RSMs of subjective valence based on the NN model and individuals RSMs of startle (Mean = .004, *t*(398)=1.47, p = .14, BF_{10} = 0.16) whereas individual RSMs of subjective valence based on

the AK model were positively related to individual RSMs of startle (Mean = .041, *t*(398)=7.32, p <.001, BF_{10} >100). Regression analysis did confirm these findings (NN model: Mean = -.004, t(398)=-1.23, p = .22, BF_{10} = 0.12; AK model: Mean = .042, *t*(398)= 7.59, p <.001, BF_{10} >100; Figure 2H). Moreover, the association between individual RSMs of startle and individual RSM of subjective valence based on the AK model remained after controlling for other models, including time-based, valence- and arousal-based categorical, and valence-based dimensional models (time-based model: Mean = 0.048, t(399)= 7.63, p <.001, BF10 >100; valence-based categorical model: Mean = 0.041, t(399)=7.34, p <.001, BF10 >100; arousal-based categorical model: Mean = .042, t(399)= 7.62, p <.001, BF10>100; valence-based dimensional model: Mean = .04, t(399)= 7.25, p <.001, BF10>100; Figure 2I).

In the replication sample, we could observe in the three tasks that the averaged RSM of startle considering intraindividual variability was a better representation of the individual RSM than the averaged RSM dismissing it. However, the associations with models of subjective valence were only partly reproduced in the PSL, but not in the PPV and Imagery tasks (Suppl. S7-S9).

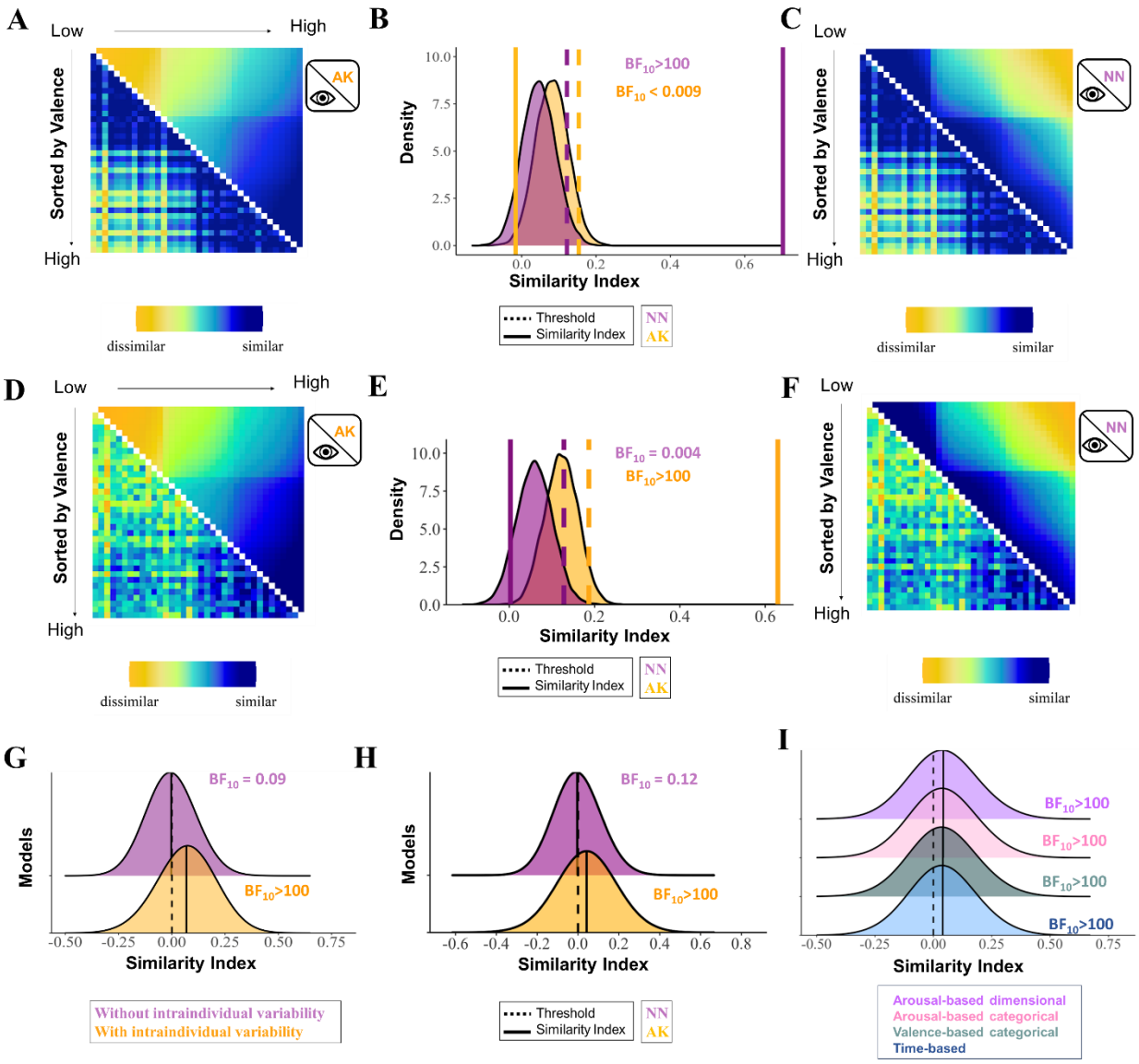

Figure 2. Association between the representational similarity matrices (RSMs) of subjective valence based on the Anna Karenina (AK) and Nearest Neighbours (NN) models (after controlling for the shared similarities), and the RSMs the startle eye blink response dismissing (A-C) and considering (D-F) intraindividual variability (discovery sample, N= 483). A) Averaged RSM of startle dismissing intraindividual variability (lower diagonal) and averaged RSM of subjective valence based on the AK model (upper diagonal). B) Results of the permutation test. C) Averaged RSM of startle dismissing intraindividual variability (lower diagonal) and averaged RSM of subjective valence based on the NN model (upper diagonal). D) Averaged RSM of startle considering intraindividual variability (lower diagonal) and averaged RSM of subjective valence based on the AK model (upper diagonal). E) Results of the permutation test. F) Averaged RSM of startle considering intraindividual variability (lower diagonal) and averaged RSM of subjective valence based on NN model (upper diagonal). G) Distribution of the association between individual RSMs of startle with the unique contribution of the averaged RSMs of startle disregarding (purple) and considering (yellow) intraindividual variability. H) Distribution of the association between individual RSMs of startle and unique contribution of the individual RSM of valence based on the AK and NN models. I) Distribution of the association between individual RSMs of startle and individual RSMs of subjective valence based on the AK model after controlling for other models (i.e., time-based, valence-based categorical, arousal-based categorical, valence-based dimensional models).

Discussion

It is widely accepted that affective experience could be conceptualized in terms of valence and arousal[1]. It is also suggested that the experience of affect is accompanied by a myriad of physiological changes. However, the correspondence between physiological responses to affective material and the experienced subjective valence and arousal is still unclear. Whereas some evidence suggests that similar affective experiences generally evoke comparable physiological changes[20,25,32] (*fingerprint* hypothesis of affect), there are also more recent studies that report within and between individual variability of physiological reactivity among events evoking similar affective experiences[14,26] (*populations* hypothesis of affect). To advance the debate on the affect-physiology relationship, we used RSA to examine the trial-by-trial similarity patterns of the subjective affective experience (i.e., valence and arousal ratings) and physiological reactivity. SCR and startle blink responses were used as physiological correlates of arousal and valence, respectively. Capitalizing on the methodological features of the RSA, we compared the *fingerprints* and *populations* hypotheses of affect by creating models of valence and arousal that align with them (NN and AK/inverted AK, respectively). Furthermore, given that the role of intraindividual variability has often not been considered, we also investigated its influence on the relationship between physiological responses and experienced affect.

Regarding the arousal dimension, models that considered vs disregarded intraindividual variability in SCR were differentially associated with NN and inverted AK models of arousal in the discovery sample. When intraindividual variability was disregarded (RSA based on response averages across the sample), we found decisive evidence for an association between the averaged RSM of SCR and the averaged RSM of arousal based on the NN model, whereas when intraindividual variability was considered (RSA based on the individual response), we

observed decisive evidence for a relationship between the averaged RSM of SCR and the averaged RMS of arousal based on the inverted AK model. More importantly, follow-up analysis revealed that models of SCR that consider intraindividual variability represent the patterns of SCR across individuals more accurately than models dismissing it. Crucially, the results could be reproduced in all three affect-inducing tasks of the replication sample (Suppl. S4-S6), indicating that these findings are reliably observed regardless of the sample and task.

Results on the correspondence between the valence dimension and the startle eye blink response showed a similar pattern to arousal-SCR relationship. In the discovery sample, the averaged RSM of startle dismissing intraindividual variability was strongly associated with the averaged RSM of valence based on the NN model, whereas the averaged RSM of startle that considered intraindividual variability showed a strong correspondence with averaged RSM of valence based on the AK model. Subsequent analysis further revealed that the RSM of startle that considered intraindividual variability offered a more prototypical representation of individual startle response patterns across individuals. Although the association between models of valence and startle responses could not be confirmed in the replication sample, we observed in all three tasks that RSMs of startle that consider intraindividual variability offer a more representative depiction across individuals than RSMs that disregard it (Figures S6-S9 G).

These findings have two main implications for the relationship between physiological responses and experienced affect. First, they demonstrate the need to consider intraindividual variability across physiological responses to gain a more valid depiction of the physiological patterns that are evoked during affect-inducing tasks. To ensure that a specific physiological response (e.g., increase in SCR) is a readout measure of an ongoing psychological process (e.g., high levels of arousal) in a particular context (affect-inducing task), two important criteria need

to be met[59]: (1) responses need to be specific, that is they should not occur if the psychological state does not take place and (2) they need to be reliable, that is they should often occur when the psychological state is given. If intraindividual variability is dismissed, the reliability and specificity of the responses remains unknown, which may lead to false conclusions and ecological fallacy[54,56]. Here, we showed that disregarding vs considering this variability leads to somewhat different physiological patterns. Importantly, across physiological measures (SCR and startle), samples (discovery and replication samples), and tasks (PPV, PSL and Imagery tasks), we found compelling evidence for models that take intraindividual variability into account to depict a more prototypical (i.e., more often shared across participants) representation of the physiological patterns across participants. These findings provide empirical evidence that the timing of averaging within the analytical process is crucial in determining the extent of meaningful variance loss. This critical step should be meticulously reported in future studies, and addressed in established guidelines. Our results further align with findings, showing the importance of person- and context-related variations to understand the relation between physiology and reported emotional experiences [26,56,60].

Our results also display that dismissing vs considering intraindividual variability may lead to somewhat opposing interpretations in relation to experienced affect. When intraindividual variability was disregarded, the averaged RSMs of SCR and startle resembled models of arousal and valence that agree with the *fingerprint* hypothesis of affect whereas averaged RSMs that took into account intraindividual variability corresponded with models of arousal and valence that assume variation are in line with the *populations* hypothesis. This difference may help understand the disparity of results observed in the literature. During the last decades, the field of affective processing has been led by proposals that concur with the *fingerprints* hypothesis of affect[23]. These proposals assume that events that are rated similar in affective terms should produce similar physiological responses. Probably driven by this

assumption, many of the previous studies on the affect-physiology relationship have opted for grouping responses from events that produced comparable affective feelings, not much focusing on intraindividual variability[29]. Consequently, many studies have shown that averaged physiological responses may covary with averaged affective reports, supporting the *fingerprint* hypothesis. However, more recent studies using data-driven multivariate methods that consider trial-by-trial relationships[14,26,56] and thus account for intraindividual variability also indicate physiological variation among similar affective experiences, in line with the *populations* hypothesis.

Second, our findings demonstrate that the association between physiological responses and experienced affect is not harmonious, but discordant. The RSMs of physiology that provide a better depiction of the physiological patterns across participants revealed a similar pattern to models of affect that assume variation, in line with the *populations* hypothesis. More specifically, in the arousal dimension results showed that trials rated with low arousal levels, typically showing low physiological responsivity (i.e., lower or no SCR), produce comparable SCR. In contrast, in trials rated with high arousal levels, which are generally accompanied by an increase in physiological responsivity, the evoked SCRs were more dissimilar to each other. An identical pattern was observed for the valence dimension, as trials higher in valence (associated with lower startle responsivity) showed comparable startle responses, while trials rated as lower in valence (i.e., more unpleasant) produced more dissimilar startle responses. These converging patterns indicate that the correspondence between physiological responses and subjective affective experience only emerge among trials generally evoking lower physiological responses, which may be due to floor effects (i.e., responses are too low to evidence differentiation). However, trials typically related to increase physiological responsivity which are those high in arousal and low in valence for SCR and startle,

respectively, are characterized by higher degrees of dissimilarity, indicating variation in the evoked physiological responses.

The present findings reflecting physiological variability among trials with similar affective experiences (particularly high arousing and low valence trials) are in line with theoretical views that concur with the principles of predictive processing[16,46,61–65]. These proposals postulate that physiological reactions associated with an incoming event do not reflect an isolated response, but depend on the physiological demands of the current situation and the interplay between the continuous physiological adjustments and the predictability of the encountered event[91,92]. These proposals further suggest that these physiological reactions are fed back to the brain by means of interoceptive mechanisms to keep the internal predictive models in check and update them, if necessary[46,61,66,67]. Importantly, these bottom-up signals (together with the myriad of sensory inputs) are incorporated as part of hierarchical, multidimensional abstractions to categorize the current state in affective terms. These proposals do not undermine the importance of physiological changes in the construction of conscious affective states but invite reframing the way the association between affect and physiological reactivity is currently outlined.

Previous studies have aimed at identifying physiological response patterns that could be indicative of a particular affective experience[4,29] and whether variations in these physiological correlates could serve as individual predictors of psychopathological vulnerability[8,13,68,69]. Here, we demonstrated that events evoking similar affective experiences produce dissimilar physiological responses (particularly in high arousing and low valence trials), suggesting that physiological reactions do not accurately reflect an affective state (e.g.,[43]). Hence, rather than focusing on deciphering physiological patterns associated with specific affective states, future research may focus on understanding the mechanisms through which situated physiological changes are used to create conscious affective experiences and/or whether the disruption of

these mechanisms may be a vulnerability factor for psychopathological disorders[70–72], as recently demonstrated that affective disorders may be characterized by a divergence between autonomic responses and subjective emotional experiences[93].

While our findings highlight the importance of considering intraindividual variability when assessing the relationship between physiological reactivity and affective experiences, shedding light on the discordant association between the dimensions of valence and arousal and physiological responses, it is important to acknowledge some limitations. In both the discovery and replication samples, the physiological reactivity and affective experience were not recorded at the same time (ratings were obtained separately after the presentation of the stimuli). Although affective experience to the same event may vary from moment to moment[73–75], the fact that our pattern of results was consistently observed after controlling for other influential factors, including time (e.g., Figures 1I and 2I), suggests that this methodological caveat may not have a strong impact on the results. Nevertheless, future studies, using designs that simultaneously track affective experience and physiological reactivity[26,76,77] may further provide valuable insights into the relationship between affective experience and physiological reactivity. The pattern of results observed for the SCR were replicated across samples and tasks but the association between startle and valence ratings based on the AK model found in the discovery sample was not replicated in the replication sample. The lack of replicability may be explained by differences in methodological characteristics. For instance, the discovery sample underwent only a single task, whereas the replication sample underwent three tasks, successively. Both studies also differed on the intensity of the startle probes (103 dB in discovery sample; 95 dB in replication sample). Hence, it could be speculated that longer durations of the experiment and/low lower startle probe intensities in the replication sample may have added variance to the physiological reactivity which may have shadowed the modulatory effects of valence (Suppl. S1).

In summary, this study validates two new RSA distance models (AK/ inverted AK) for peripheral psychophysiological data and highlights the necessity of considering intraindividual variability with multivariate approaches to create more representative physiological patterns across participant. This step is particularly crucial as we could demonstrate that considering vs. dismissing intraindividual variability may lead to somewhat opposing conclusions in the affect-physiology relationship. Furthermore, our results highlight a discordance between physiological responses and subjective affective experience, as suggested by the fact that events evoking similar affective experiences may produce dissimilar physiological responses. These findings invite to reassess the notion of physiological reactivity as a potential indicator of affective states and prompt a shift towards understanding the mechanisms through which physiological changes contribute to conscious affective experiences.

Materials and Methods

Participants

Discovery Sample.

The sample comprises 685 individuals (424 female, 246 male; M_{age} = 25.36, SD_{age} = 5.57, from four sub-studies (for details[13]), which were conducted at the University Medical Center Hamburg-Eppendorf. Inclusion criteria encompassed physically and mentally healthy individuals between the ages of 18 and 50 who self-identified as of Caucasian origin. Exclusion criteria encompassed medication usage (except for oral contraceptives), a history of psychiatric diagnoses, and pregnancy. All participants had normal or corrected-to-normal vision and provided written informed consent in compliance with the protocol approved by the Ethics Committee of the General Medical Council Hamburg (PV 2755), adhering to the principles of the Declaration of Helsinki. Results on a different research question based on this dataset were published previously[13].

Replication Sample.

A total of sixty-four healthy students (51 women, 13 men; M_{age} = 23.43, SD_{age} = 4.11) from the University of Potsdam took part in the study in exchange of course credits or financial compensation. All participants had normal or corrected-to-normal vision. Each individual provided written informed consent for a protocol approved by the ethics committee of the University of Potsdam (28/2018). Before the experimental session, participants were screened and only invited to participate if they did not fulfil any of the following exclusion criteria: neurological or mental disorders, brain surgery, undergoing chronic or acute medication, history of migraine and/or epilepsy. The extended results from the replication sample were published somewhere else[14].

Material, design and procedure

Discovery Sample.

The procedure of the passive picture viewing (PPV) task has been described previously elsewhere [13,33]. In brief, the task consisted of two main phases: an initial 8-trial startle habituation phase (see below), followed by passive viewing of a set of 36 images. These images were selected from both the *International Affective Picture System* (IAPS)[15] and *EmoPicS*[78] databases, with an equal distribution of twelve images per emotional valence (negative, neutral, positive; detailed information can be found in[13,33]. The 36 images were presented twice, resulting in a total of 72 trials.

For positive valence, erotic images were tailored to match the subject's sexual preference (i.e., same-sex or opposite-sex pictures), while negative images were selected to depict personally threatening situations (e.g., a gun pointed at the subject or an animal attacking the subject). Each image was displayed for 6 s, separated by randomized inter-trial intervals (ITIs) of 10, 12, or 14 s during which a fixation cross was presented. In 50% of the picture presentations, acoustic startle probes (103 dB, white noise) were presented either 4.5 or 5.5 s after the onset of the stimulus. To prevent predictability of the startle probes, 12 startle probes were distributed across the long ITIs (i.e., 14s), presented 8 s after ITIs onset.

During the initial startle habituation phase, eight startle probes (ITI: 11s) were presented while a fixation cross was displayed. Additionally, to maintain the subjects' alertness, three occasional oddball trials were incorporated, as detailed in[33].

Post-experimentally, participants rated the previously seen pictures in terms of valence and arousal, using a computerized version of the self-assessment manikin (SAM[6]).

Replication Sample.

The experimental session comprised five blocks. First, participants' physiological responses were measured at rest for 6 minutes. The following three sections consisted of affect-inducing tasks, whose order was counterbalanced across participants, with different stimulus material: a passive picture viewing task (PPV), a passive sound listening task (PSL), and an imagery task. In the last block of the experimental session, participants rated the stimuli used in the three prior tasks in terms of valence and arousal, using a computerized version of the self-assessment manikin (SAM[6]).

Passive Picture Viewing (PPV) and Passive Sound Listening (PSL) tasks. A total of 36 images from the IAPS[15] and 36 sounds from the *International Affective Digital Sounds* (IADS[79]) were selected based on their normative valence and arousal ratings and grouped in three affective categories (pleasant, unpleasant, and neutral). Each stimulus (image or sound) was followed by an ITI of 14, 16, or 18 s. In half of the trials, an acoustic startle probe (95 dB, white noise) was presented 5.5 or 6.5 s after cue onset. In half of the 18-second-long ITIs, the startle probe was also presented 9 s after cue offset (i.e., 18 startle probes during picture viewing/sound listening and 6 probes during ITIs). To account for startle habituation effects, two startle probes were presented one minute before each experimental block.

Imagery task. Eighteen scripts were selected from different sources: the *Affective Norms for English Text* (ANET; N = 15[15]), from prior studies[80], and from in-house sets. All scripts were classified in three affective categories based on their normative valence and arousal ratings (for ANET scripts). Each script was presented twice, resulting in a total of 36 trials. Participants were instructed to read the scripts during presentation and, after script offset, to imagine themselves in the situation as vividly as possible. During the imagery period a circle was presented in the middle of the screen. As for the PPV and PSL tasks, the presentation order of the affective categories was pseudorandomized. The order of the scripts within each category was also pseudorandomized in 8 different orders. Scripts were presented for 12 s and followed

by a 9-second imagery time window. The ITI varied between 14, 16, or 18 s. In 18 trials (once for each script), the acoustic startle probe was presented 5.5 or 6.5 s after onset of the imagery. Furthermore, six startle probes were presented 9 s after imagery cue offset in half of the 18-second-long ITIs. To account for startle habituation effects, two startle probes were presented one minute before the experimental block began.

Affective Rating task. The subjective valence and arousal ratings for each stimulus (36 images, 36 sounds or the 18 imagery scripts) were collected using the online platform *soscisurvey* (soscisurvey.de; [54]). Which stimulus modality was rated first was dependent on the order of the tasks in the previous three blocks (e.g., if the participant underwent the PPV in block 2, the Imagery task in block 3 and the PSL in block 4, then images were rated first, then imagery scripts, and lastly sounds). Within each stimulus-set, the presentation of the stimuli was randomized. Participants were instructed to indicate on a SAM scale (1-9, see above) how they felt while viewing/listening/imagining the stimuli in terms of arousal and valence.

Data recording, reduction and response definition.

Discovery Sample.

The recording of startle eyeblink responses and skin conductance responses (SCR) was conducted using a BIOPAC MP100 amplifier manufactured by BIOPAC Systems Inc. The data acquisition was managed using AcqKnowledge software version 3.9.2.

SCRs were assessed using self-adhesive Ag/AgCl electrodes, each measuring 55 millimeters in length. These electrodes were attached to the palmar side of the left hand (distal and proximal hypothenar regions). Prior to electrode placement, the hands were washed with warm water without the use of soap. SCRs were continuously recorded at a sampling rate of 1000 Hz and a gain of 5 micro-ohms (μΩ).

Subsequently, the recorded data were down-sampled offline to a rate of 10 Hz. SCR amplitudes, measured in microsiemens (μS), were then scored semi-automatically utilizing a custom-made program. The largest SCR response with an onset latency between 0.9 and 4.0 s after the onset of the stimulus was identified, as recommended by published guidelines[81].

Fear-potentiated startle responses were recorded beneath the right eye through the use of two Ag/AgCl electromyogram (EMG) electrodes strategically positioned over the orbicularis oculi muscle, along with one electrode placed on the participant's forehead to serve as a reference point. The startle data were collected with a gain setting of 5000 at a rate of 1000 Hz. Online, the data underwent band-pass filtering within the range of 28 to 500 Hz, followed by rectification and integration (averaging over 20 samples).

The subsequent data analysis involved manual scoring based on a "foot to peak" approach, focusing on the time window from 20 to 150 ms following the onset of the startle probe. This scoring process adhered to established guidelines[82] and was executed using the same program utilized for skin conductance responses (SCRs). Any responses affected by a blink occurring within 50 ms before the administration of the startle probe, recording anomalies, or excessive baseline activity (as determined by visual inspection) were categorized as missing values.

Responses that were compromised by recording anomalies, such as electrode detachment or responses extending beyond the defined sampling window, were also excluded and considered as missing data points.

Replication Sample.

Skin Conductance Response (SCR). An MP-160 BIOPAC system (BIOPAC systems, Goleta, CA) with 2 shielded 8 mm Ag-AgCl sensors applied on the thenar and hypothenar

eminence of the palmar surface of the participant's left hand was used to collect electrodermal activity. SCRs were extracted using Continuous Decomposition Analysis in Ledalab Version 3.4.9[83]. The skin conductance data, initially recorded at a 2000 Hz sampling rate, were down sampled to a resolution of 50 Hz and parameters of the impulse response function (IRF) used to decompose the SC were optimized using four sets of initial values (see for details on Continuous Decomposition Analysis[83]). For the PPV and PSL task, the response window was set from 1 to 4 s after stimulus onset. For the Imagery task, the response window was set from 1 to 4 sec after the beginning of the imagery part. Both time-series and the averaged phasic driver[83] were extracted as dependent variables. To ensure normalization of the data, the averaged phasic driver was log transformed (log[1+SCR driver]). Prior to performing the analysis, a visual inspection of the data was carried out to detect possible anomalies during recording. Due to anomalies during recording (excessive artifacts or near-to-zero skin conductance levels) 5 participants were excluded from the PPV task, 3 from the PSL task, and 3 from the Imagery task.

Startle Eyeblink Response. To measure the eyeblink component of the startle response, a MP-160 BIOPAC system (BIOPAC systems, Goleta, CA) was employed to record electromyographic (EMG) activity at sampling rate of 2000 Hz over the left orbicularis oculi, using two Ag/AgCl miniature surface electrodes. To evoke the startle blink reflex, an acoustic startle probe, which consisted of 95 dB, 50 ms burst of white noise with a near instantaneous rise time, was used. The sound was administered in both ears using Audio Technica ATH-PRO700 MK2 headphones. Following EMG startle guidelines[82], the data were filtered (high-pass filter of 60 Hz), rectified, and smoothed with a time constant of 10 ms. Startle blink responses were defined using in-house scripts in MATLAB 2020a. All trials were visually inspected and manually scored (c.f., guidelines for startle studies[82]). Startle responses were considered when starting 20-120 ms after probe onset and peaking within 150 ms. The

magnitude of the eye blink response (in microvolts) was measured from onset to peak. Trials with no detectable blinks were scored as zero responses, and trials with excessive baseline activity or recording artifacts were scored as missing. Participants with more than 50% of non-responses[55] were excluded from the analysis (3 excluded participants from the PPV task, 3 from the PSL task, and 2 from the Imagery task).

Representational similarity analysis (RSA)

Computation of Representational Similarity Matrices (RSMs) of models based on affective subjective experience.

We defined three models of subjective experience of valence and arousal: the Nearest Neighbours (NN) model, the Anna Karenina (AK[58]) model and its inverted form (inverted AK). The NN assumes that a response on a specific trial is similar to responses on dimensionally close trials (e.g., low arousing trials should evoke similar responses) [58], implying that events evoking similar affective experiences produce comparable physiological responses in line with the *fingerprint* hypothesis. For instance, the NN model predicts similar SCRs among low arousing trials and among high arousing trials, separately, but dissimilar SCRs between high and low arousing trials. The AK-related models assume similarities between dimensionally close trials on one side of a dimension, but dissimilarities between dimensionally close trials on the other side of the dimension[58]. More concretely, the AK model represents higher similarity between trials located in the higher extreme of a dimension but higher dissimilarity between trials from the lower end of the dimension (e.g., similar physiological responding across high arousing/valence trials, but dissimilar physiological reactions across low arousing/valence trials). The inverted AK model represents the opposite pattern (e.g., similar physiological responding across low arousing/valence trials, but dissimilar physiological reactions across high arousing/valence trials). Given that the AK and inverted AK models are

complimentary (they are perfectly, negatively related), to facilitate the interpretation of the results, we first visually inspected the physiological RSMs and determined the usage of the model which could resemble them most (i.e., showed more similarity rather than dissimilarity).

We constructed NN and AK-related models for arousal and valence ratings, using a two-step procedure. First, an RSM for each participant was constructed by sorting the trials based on the reported subjective experience (i.e., valence or arousal ratings, e.g., Jin et al., 2015) and calculating the trial-by-trial similarity scores. The formulae used for each model are as follows:

NN Model: $Similarity = V_{max} - \left(\left|\left(V_{Ti} - V_{Tj}\right)\right|\right)$

AK Model: $Similarity = mean\left(V_{Ti,}V_{Tj}\right)/N_{variables}$

Inverted AK Model: $Similarity = 1/\left(\frac{mean\left(V_{Ti,}V_{Ti+1}\right)}{N_{variables}}\right)$

V_{max} represents the maximum value of the dataset, and it is used as a constant to ensure that all similarity values stay in the positive range. V_{ti} represents the score of trial *i*. V_{tj} denotes the score of trial *j*. $N_{variables}$ indicates the number of trials (e.g., 36 or 18 trials).

In a second step, an averaged RSM was calculated by computing the average of the RSM across participants (so-called ‘averaged RSM’, see Box for terminology).

Computation of the physiological RSMs.

SCR. As a dependent variable to calculate the RSMs, the SCR amplitude (log transformed to achieve normal distribution) was used in the discovery sample and the log transformed mean phasic driver (as an amplitude approximate derived from *Ledalab*[83]) in the replication sample. This approach for extracting the SCR may vary from our previous work, where the phasic driver vector was used (see Suppl. S2, S3 for additional comparative analyses).

Startle eyeblink response. To construct the RSMs of the startle, t-transformed magnitudes were used as dependent variables as suggested in the guidelines (Blumenthal et al., 2005). In trials in which startle response could not be quantified (i.e., missing trials), the similarity scores could not be calculated, and were thus treated as missing values. Although this strategy resulted in an uneven number of cases per participant, it prevented us from excluding participants with missing trials from the analysis.

To test for the effects of intraindividual variation, two different types of averaged RSMs of physiological responses were calculated, one disregarding intraindividual variability, and one considering it. To construct the averaged RSMs dismissing intraindividual variability, trials were first sorted based on the subjective affective ratings (SCRs were sorted based on subjective ratings of arousal and startle responses were sorted based on subjective ratings of valence) and trials were averaged across participants. The averaged RSM was calculated based on the resulting averaged scores for each trial (N trials = 36 for SCR in both samples, except for the Imagery task [N trials = 18]; N trials = 36 for the startle in the discovery sample and N trials = 18 in the replication sample).

For the construction of the RSMs of physiological responses considering intraindividual variability, we followed an identical procedure to the construction of the RSMs of the models

of subjective affective experience. Trials were first sorted based on the individual subjective affective ratings of interest (i.e., arousal for SCR- and valence for startle-based RSMs) and the trial-by-trial similarity was calculated (i.e., using the NN formula) for each participant. This resulted in a single RSM for each individual participant. Thereafter, an averaged RSM was calculated by averaging the individual RSMs across the participants.

Computation of RSMs of alternative models.

To ensure that the convergence between the RSMs based on subjective affective experience and the physiological RSMs was beyond other factors known to influence psychophysiological responses[84] we computed RSMs based on the IAPS categorization of picture valence (pleasant vs unpleasant vs neutral) and arousal categories (low vs high arousing), and time (operationalized by the location in which trials position were presented). For the construction of the alternative RSMs, trials were initially sorted based on the subjective affective experience of interest (i.e., arousal for SCR- and valence for startle-based RSMs) and then the trial-by-trial similarities were calculated for each participant (using the NN-like formula), using as dependent variable nominal indexes for the category-based RSMs (e.g., for the valence category, pleasant images were coded as 1; unpleasant as 2; and neutral as 3; for the arousal category, high arousing were coded as 1 and low arousing as 2), and the trial number (e.g., from 1 to 36) for the time-based RSM.

Comparing model and physiological RSMs

Firstly, the two averaged physiological RSMs (without and with intraindividual variability) were compared to each other and to the averaged RSM of arousal and valence based on NN and AK models. More specifically, the averaged RSMs of subjective arousal based on the NN and inverted AK models were compared to the averaged RSMs of SCR, whereas the averaged RSMs of subjective valence based on NN and AK models were compared to the

averaged RSMs of startle. The correspondence between averaged RSM of arousal and valence and the averaged RSMs of physiological responses was tested, using Spearman's rank correlation coefficients, using the *stats* package in R. Because RSMs are mirrored along the diagonal, to avoid inflated correlations, only the values below the diagonal were considered for the correlation.

The statistical significance of the similarity between averaged RSMs of arousal and valence and the averaged RSMs of physiological responses was examined, using permutation tests by randomizing the modality labels of each pairing (i.e., model vs. physiological). The labels were iterated 10,000 times, using Monte Carlo random permutations. For each iteration, the correlation between RSMs was computed. These estimates served as a null distribution against which the actual correlation coefficient was compared. If the actual correlation coefficient was above 95% of the null distribution, we rejected the null hypothesis that the physiological and subjective RSMs are unrelated. We also used a Bayesian framework to quantify evidence in favor of the null (no positive relationship) or alternative (a positive relationship) hypotheses based on the results of the permutation test. To do so, we calculated the Bayes factor test for linear correlations, with the *BayesFactor*[85] package in R environment (v. 4.3.3), using the distribution of the permutation test as the null interval.

Although the NN and AK and inverted AK models assume a different relationship between trials along a dimension (e.g., arousal), in some ranges of the dimension, they may show concurrent similarity patterns. For instance, for arousal the NN and inverted AK models show different patterns in high arousing trials, but similar patterns in low arousing trials. This implies that if a significant correspondence between the models and the physiological RSMs is observed, it may be unclear whether it is explained by the shared variance or by the uniqueness of each model. Thus, to further elucidate the unique contribution of each of the RSM models to the convergence with physiological RSMs, we performed regression analysis (e.g., the NN

was regressed out from the inverted AK model and vice versa) and compared the correspondence between the residuals and the averaged physiological RSM[86], using permutation tests and Bayesian testing.

To assess which of the averaged physiological RSMs (without and with intraindividual variability) was a better representative of the individual RSMs, in a subsequent step, individual RSMs were compared to both averaged RSMs (without and with intraindividual variability). To do so, similarity indexes (i.e., Spearman's correlation coefficient) of the relationship between the individual RSMs and the averaged RSMs were extracted for each participant. The significance of the correspondence was tested by performing a one-sided t-test across participants, testing the null hypothesis of no correspondence (i.e., similarity does not differ from zero) between averaged and individual physiological RSMs. Bayesian testing was further performed to elucidate the evidence in favor of the null or alternative hypotheses. Given that both averaged RSMs may be related, we tested the unique contribution of each averaged RSM (i.e., without and with intraindividual variability). Similarity indexes of the relation between individual RSMs and the residuals of each of the averaged RSMs (extracted from regressing one model out of the other) were extracted for all participants and the correspondence was tested, using one-sided t-tests.

In addition, each participant's physiological RSM was compared to the NN and AK-related models of arousal and valence. This was done by extracting similarity indexes between the physiological and subjective affective experience RSMs for each participant. The significance of the correspondence was tested by performing a one-sided t-test across participants, testing the null hypothesis of no correspondence (i.e., similarity does not differ from zero) between models and physiological RSMs. Bayesian testing was further performed to elucidate the evidence in favor of the null or alternative hypotheses. The unique contribution of the tested models (i.e., NN and inverted AK for SCR and NN and AK for startle) for the

correspondence with physiological RSMs was further tested, using the abovementioned strategy on the model residuals (extracted from regressing one model out of the other). Additionally, we also tested whether the models explained variance beyond alternative models such as valence-based and arousal-based categorical models, time, and the alternative dimensional model (subjective valence for SCR and arousal valence for startle) by using the same regression procedure.

References

1. Barrett, L. F. & Russell, J. A. The Structure of Current Affect: Controversies and Emerging Consensus. *Curr. Dir. Psychol. Sci.* **8**, 10–14 (1999).

2. Cacioppo, J. T. & Berntson, G. G. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychol. Bull.* **115**, 401–423 (1994).

3. Kuppens, P., Tuerlinckx, F., Russell, J. A. & Barrett, L. F. The relation between valence and arousal in subjective experience. *Psychol. Bull.* **139**, 917–940 (2013).

4. Lang, P. J., Bradley, M. M. & Cuthbert, B. N. Emotion, attention, and the startle reflex. *Psychol. Rev.* **97**, 377–395 (1990).

5. Reisenzein, R. Pleasure-arousal theory and the intensity of emotions. *J. Pers. Soc. Psychol.* **67**, 525–539 (1994).

6. Bradley, M. M. & Lang, P. J. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **25**, 49–59 (1994).

7. Bradley, M. M. Emotional memory: A dimensional analysis. in *Emotions: Essays on emotion theory* 97–134 (Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1994).

8. Hamm, A. O. Fear, anxiety, and their disorders from the perspective of psychophysiology. *Psychophysiology* **57**, e13474 (2020).

9. Milad, M. R. & Quirk, G. J. Fear extinction as a model for translational neuroscience: ten years of progress. *Annu. Rev. Psychol.* **63**, 129–151 (2012).

10. Yancey, J. R., Vaidyanathan, U. & Patrick, C. J. Aversive startle potentiation and fear pathology: Mediating role of threat sensitivity and moderating impact of depression. *Int. J. Psychophysiol.* **98**, 262–269 (2015).

11. Vaidyanathan, U., Welo, E. J., Malone, S. M., Burwell, S. J. & Iacono, W. G. The effects of recurrent episodes of depression on startle responses. *Psychophysiology* **51**, 103–109 (2014).

12. Ribes-Guardiola, P. *et al.* Attention-affect interactions and triarchic psychopathy: New electrophysiological insights from the late positive potential. *Psychophysiology* **60**, e14222 (2023).

13. Koppold, A., Kastrinogiannis, A., Kuhn, M. & Lonsdorf, T. B. Watching with Argus eyes: Characterization of emotional and physiological responding in adults exposed to childhood maltreatment and/or recent adversity. *Psychophysiology* **60**, e14253 (2023).

14. Ventura-Bort, C., Wendt, J. & Weymar, M. New insights on the correspondence between subjective affective experience and physiological responses from representational similarity analysis. *Psychophysiology* **59**, e14088 (2022).

15. Bradley, M. M. & Lang, P. J. Emotion and motivation. in *Handbook of psychophysiology, 3rd ed* 581–607 (Cambridge University Press, New York, NY, US, 2007). doi:10.1017/CBO9780511546396.025.

16. Barrett, L. F. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* nsw154 (2016) doi:10.1093/scan/nsw154.

17. Friedman, B. H. Feelings and the body: The Jamesian perspective on autonomic specificity of emotion. *Biol. Psychol.* **84**, 383–393 (2010).

18. Lang, P. J., Greenwald, M. K., Bradley, M. M. & Hamm, A. O. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* **30**, 261–273 (1993).

19. vanOYEN Witvliet, C. & Vrana, S. R. Psychophysiological responses as indices of affective dimensions. *Psychophysiology* **32**, 436–443 (1995).

20. Lang, P. J. & Bradley, M. M. Emotion and the motivational brain. *Biol. Psychol.* **84**, 437–450 (2010).

21. Roelofs, K. & Dayan, P. Freezing revisited: coordinated autonomic and central optimization of threat coping. *Nat. Rev. Neurosci.* **23**, 568–580 (2022).

22. Lang, P. J. Emotion's Response Patterns: The Brain and the Autonomic Nervous System. *Emot. Rev.* **6**, 93–99 (2014).

23. Lang, P. J. The varieties of emotional experience: a meditation on James-Lange theory. *Psychol. Rev.* **101**, 211–221 (1994).

24. Levenson, R. W. The Autonomic Nervous System and Emotion. *Emot. Rev.* **6**, 100–112 (2014).

25. Hamm, A. O., Greenwald, M. K., Bradley, M. M. & Lang, P. J. Emotional learning, hedonic change, and the startle probe. *J. Abnorm. Psychol.* **102**, 453–465 (1993).

26. Hoemann, K. *et al.* Context-aware experience sampling reveals the scale of variation in affective experience. *Sci. Rep.* **10**, 12459 (2020).

27. Siegel, E. H. *et al.* Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychol. Bull.* **144**, 343–393 (2018).

28. Ekman, P. & Cordaro, D. What is meant by calling emotions basic. *Emot. Rev.* **3**, 364–370 (2011).

29. Bradley, M. M., Codispoti, M., Cuthbert, B. N. & Lang, P. J. Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion* **1**, 276–298 (2001).

30. Walker, D. L., Cassella, J. V., Lee, Y., De Lima, T. C. & Davis, M. Opposing roles of the amygdala and dorsolateral periaqueductal gray in fear-potentiated startle. *Neurosci. Biobehav. Rev.* **21**, 743–753 (1997).

31. Terburg, D. *et al.* The Basolateral Amygdala Is Essential for Rapid Escape: A Human and Rodent Study. *Cell* **175**, 723-735.e16 (2018).

32. Lang, P. J., Greenwald, M. K., Bradley, M. M. & Hamm, A. O. Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology* **30**, 261–273 (1993).

33. Kuhn, M. *et al.* The Neurofunctional Basis of Affective Startle Modulation in Humans: Evidence From Combined Facial Electromyography and Functional Magnetic Resonance Imaging. *Biol. Psychiatry* **87**, 548–558 (2020).

34. Bradley, M. M. & Lang, P. J. Affective reactions to acoustic stimuli. *Psychophysiology* **37**, 204–215 (2000).

35. Brouwer, A.-M., Van Wouwe, N., Mühl, C., Van Erp, J. & Toet, A. Perceiving blocks of emotional pictures and sounds: effects on physiological variables. *Front. Hum. Neurosci.* **7**, (2013).

36. Vrana, S. R., Spence, E. L. & Lang, P. J. The startle probe response: A new measure of emotion? *J. Abnorm. Psychol.* **97**, 487–491 (1988).

37. Bradley, M. M., Codispoti, M. & Lang, P. J. A multi-process account of startle modulation during affective perception. *Psychophysiology* **43**, 486–497 (2006).

38. Cuthbert, B. N., Bradley, M. M. & Lang, P. J. Probing picture perception: activation and emotion. *Psychophysiology* **33**, 103–111 (1996).

39. Lang, P. J. The emotion probe. Studies of motivation and attention. *Am. Psychol.* **50**, 372–385 (1995).

40. Guerra, P., Sánchez-Adam, A., Anllo-Vento, L., Ramírez, I. & Vila, J. Viewing Loved Faces Inhibits Defense Reactions: A Health-Promotion Mechanism? *PLoS ONE* **7**, e41631 (2012).

41. WannemÃ¼ller, A., Sartory, G., Elsesser, K., Lohrmann, T. & JÃ¶hren, H. P. Modality of fear cues affects acoustic startle potentiation but not heart-rate response in patients with dental phobia. *Front. Psychol.* **6**, (2015).

42. Roy, M., Mailhot, J.-P., Gosselin, N., Paquette, S. & Peretz, I. Modulation of the startle reflex by pleasant and unpleasant music. *Int. J. Psychophysiol.* **71**, 37–42 (2009).

43. Joffily, M. & Coricelli, G. Emotional Valence and the Free-Energy Principle. *PLoS Comput. Biol.* **9**, e1003094 (2013).

44. Hesp, C. *et al.* Deeply Felt Affect: The Emergence of Valence in Deep Active Inference. *Neural Comput.* **33**, 398–446 (2021).

45. Lee, K. M., Ferreira-Santos, F. & Satpute, A. B. Predictive processing models and affective neuroscience. *Neurosci. Biobehav. Rev.* **131**, 211–228 (2021).

46. Sennesh, E. *et al.* Interoception as modeling, allostasis as control. *Biol. Psychol.* **167**, 108242 (2022).

47. Schulkin, J. & Sterling, P. Allostasis: A Brain-Centered, Predictive Mode of Physiological Regulation. *Trends Neurosci.* **42**, 740–752 (2019).

48. Chikazoe, J., Lee, D. H., Kriegeskorte, N. & Anderson, A. K. Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* **17**, 1114–1122 (2014).

49. Jin, J., Zelano, C., Gottfried, J. A. & Mohanty, A. Human Amygdala Represents the Complete Spectrum of Subjective Valence. *J. Neurosci.* **35**, 15145–15156 (2015).

50. Skerry, A. E. & Saxe, R. Neural Representations of Emotion Are Organized around Abstract Event Features. *Curr. Biol.* **25**, 1945–1954 (2015).

51. Cacioppo, J. T. *et al.* Relationship between facial expressiveness and sympathetic activation in emotion: A critical review, with emphasis on modeling underlying mechanisms and individual differences. *J. Pers. Soc. Psychol.* **62**, 110–128 (1992).

52. Posner, J., Russell, J. A. & Peterson, B. S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **17**, 715–734 (2005).

53. Hamm, A. O., Cuthbert, B. N., Globisch, J. & Vaitl, D. Fear and the startle reflex: Blink modulation and autonomic response patterns in animal and mutilation fearful subjects. *Psychophysiology* **34**, 97–107 (1997).

54. Freedman, D. A. Ecological inference and the ecological fallacy. in *International encyclopedia of the social & behavioral sciences* vol. Vol. 549 4027–4030 (2001).

55. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* **11**, 126 (2013).

56. McVeigh, K., Kleckner, I. R., Quigley, K. S. & Satpute, A. B. Fear-related psychophysiological patterns are situation and individual dependent: A Bayesian model comparison approach. *Emot. Wash. DC* **24**, 506–521 (2024).

57. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* (2008) doi:10.3389/neuro.06.004.2008.

58. Finn, E. S. *et al.* Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage* **215**, 116828 (2020).

59. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements - Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, Seth D. Pollak, 2019. https://journals.sagepub.com/doi/10.1177/1529100619832930.

60. Toner, E. R. *et al.* Wearable Sensor-based Multimodal Physiological Responses of Socially Anxious Individuals across Social Contexts. Preprint at https://doi.org/10.48550/arXiv.2304.01293 (2023).

61. Barrett, L. F. & Simmons, W. K. Interoceptive predictions in the brain. *Nat. Rev. Neurosci.* **16**, 419–429 (2015).

62. Katsumi, Y. *et al.* Correspondence of functional connectivity gradients across human isocortex, cerebellum, and hippocampus. *Commun. Biol.* **6**, 401 (2023).

63. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).

64. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).

65. Hutchinson, J. B. & Barrett, L. F. The power of predictions: An emerging paradigm for psychological research. *Curr. Dir. Psychol. Sci.* **28**, 280–291 (2019).

66. Seth, A. K. Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* **17**, 565–573 (2013).

67. Khalsa, S. S. *et al.* Interoception and Mental Health: A Roadmap. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 501–513 (2018).

68. Duits, P. *et al.* Updated Meta-Analysis of Classical Fear Conditioning in the Anxiety Disorders. *Depress. Anxiety* **32**, 239–253 (2015).

69. Beckers, T. *et al.* Understanding clinical fear and anxiety through the lens of human fear conditioning. *Nat. Rev. Psychol.* **2**, 233–245 (2023).

70. Smith, R. *et al.* A Bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *PLOS Comput. Biol.* **16**, e1008484 (2020).

71. Smith, R. *et al.* Perceptual insensitivity to the modulation of interoceptive signals in depression, anxiety, and substance use disorders. *Sci. Rep.* **11**, 2108 (2021).

72. Brown, C. L. *et al.* Coherence between subjective experience and physiology in emotion: Individual differences and implications for well-being. *Emot. Wash. DC* **20**, 818–829 (2020).

73. Baskin-Sommers, A. R., Curtin, J. J. & Newman, J. P. Emotion-modulated startle in psychopathy: Clarifying familiar effects. *J. Abnorm. Psychol.* **122**, 458–468 (2013).

74. Asutay, E. *et al.* Affective calculus: The construction of affect through information integration over time. *Emotion* **21**, 159–174 (2021).

75. Strube, A., Rose, M., Fazeli, S. & Büchel, C. Alpha-to-beta- and gamma-band activity reflect predictive coding in affective visual processing. *Sci. Rep.* **11**, 23492 (2021).

76. Kragel, P. A. & Labar, K. S. Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. *Emot. Wash. DC* **13**, 681–690 (2013).

77. Eisenbarth, H. *et al.* Emotional arousal pattern (EMAP): A new database for modeling momentary subjective and psychophysiological responding to affective stimuli. *Psychophysiology* e14446 (2023) doi:10.1111/psyp.14446.

78. Wessa, M. *et al.* EmoPics: Subjektive und psychophysiologische Evaluation neuen Bildmaterials für die klinisch-biopsychologische Forschung. *Z. Für Klin. Psychol. Psychother.* **39**, (2010).

79. Bradley, M. *The International Affective Digitized Sounds (IADS): Stimuli, Instruction Manual and Affective Ratings*. (NIMH Center for the Study of Emotion and Attention, 1999).

80. Limberg, A., Barnow, S., Freyberger, H. J. & Hamm, A. O. Emotional vulnerability in borderline personality disorder is cue specific and modulated by traumatization. *Biol. Psychiatry* **69**, 574–582 (2011).

81. Boucsein, W. *Electrodermal Activity, 2nd Ed.* xviii, 618 (Springer Science + Business Media, New York, NY, US, 2012). doi:10.1007/978-1-4614-1126-0.

82. Blumenthal, T. D. *et al.* Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology* **42**, 1–15 (2005).

83. Benedek, M. & Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**, 80–91 (2010).

84. Boucsein, W. *et al.* Publication recommendations for electrodermal measurements. *Psychophysiology* **49**, 1017–1034 (2012).

85. Morey, R. D. *et al.* BayesFactor: Computation of Bayes Factors for Common Designs. (2024).

86. Mitchell, D. J., Beckmann, C. & Biro, P. A. Understanding the unexplained: The magnitude and correlates of individual differences in residual variance. *Ecol. Evol.* **11**, 7201 (2021).

87. Schiller, D., Yu, A. N. C., Alia-Klein, N., Becker, S., Cromwell, H. C., Dolcos, F., Eslinger, P. J., Frewen, P., Kemp, A. H., Pace-Schott, E. F., Raber, J., Silton, R. L., Stefanova, E., Williams, J. H. G., Abe, N., Aghajani, M., Albrecht, F., Alexander, R., Anders, S., … Lowe, L. (2024). The Human Affectome. *Neuroscience & Biobehavioral Reviews*, 158, 105450. https://doi.org/10.1016/j.neubiorev.2023.105450

88. Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160007. https://doi.org/10.1098/rstb.2016.0007

89. Ferreira-Santos, F. (2016). The role of arousal in predictive coding. *Behavioral and Brain Sciences*, 39, e207. https://doi.org/10.1017/S0140525X15001788

90. Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, 7(3), 249–253. https://doi.org/10.1177/014662168300700301

91. Konu, D., Mckeown, B., Turnbull, A., Siu Ping Ho, N., Karapanagiotidis, T., Vanderwal, T., McCall, C., Tipper, S. P., Jefferies, E., & Smallwood, J. (2021). Exploring patterns of ongoing thought under naturalistic and conventional task-based conditions. *Consciousness and Cognition*, 93, 103139. https://doi.org/10.1016/j.concog.2021.103139

92. Smallwood, J., Turnbull, A., Wang, H., Ho, N. S. P., Poerio, G. L., Karapanagiotidis, T., Konu, D., Mckeown, B., Zhang, M., Murphy, C., Vatansever, D., Bzdok, D., Konishi, M., Leech, R., Seli, P., Schooler, J. W., Bernhardt, B., Margulies, D. S., & Jefferies, E. (2021). The neural correlates of ongoing conscious thought. *iScience*, 24(3), 102132. https://doi.org/10.1016/j.isci.2021.102132

93. Cuve, H. C. J., Harper, J., Catmur, C., & Bird, G. (2023). Coherence and divergence in autonomic-subjective affective space. *Psychophysiology*, 60(6), e14262. https://doi.org/10.1111/psyp.14262