Quantification choices for individual differences: An example of mapping self-report to psychophysiological responses

Jayne Morriss[1*], Nicolo Biagi[2], Shannon Wake [3]

[1]School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, UK

[2]Henley Business School, Business Informatics Systems and Accounting, Informatics Research Centre, University of Reading, UK

[3]School of Psychology, University of Reading, UK

* Correspondence:

Jayne Morriss

School of Psychology

B44 University Rd

University of Southampton

Southampton

SO17 1PS

j.morriss@soton.ac.uk

Abstract

A popular focus in affective neuroscience research has been to map the relationships between individual differences (e.g. personality and environmental experiences) and psychophysiological responses, in order to further understand the effect of individual differences upon neurobehavioral systems that support affect and arousal. Despite this trend, there have been a lack of practical examples demonstrating how the quantification of individual differences (e.g. categorical or continuous) impacts the observed relationships between different units of analysis (e.g. self-report > psychophysiological responses). To address this gap, we conducted a two-stage aggregated meta-analysis of self-reported intolerance of uncertainty (IU) and skin conductance responses during threat extinction (k = 18, n = 1006) using different quantification choices for individual differences in self-reported intolerance of uncertainty (continuous, categorical via median split, and categorical via extremes – one standard deviation above/below). Results from the meta-analyses revealed that the different quantification techniques produced some consistent (e.g. higher IU was significantly associated with skin conductance responding during late extinction training) and inconsistent IU-related effects. Furthermore, the number of statistically significant effects and effect sizes varied based on the quantification of individual differences in IU (e.g. categorical, compared to continuous was associated with more statistically significant effects, and larger effect sizes). The current study highlights how conducting different quantification methods for individual differences may help researchers understand the individual difference construct of interest (e.g. characterisation,

measurement), as well as examine the stability and reliability of individual difference-based effects and correspondence between various units of analysis.

Keywords: Intolerance of Uncertainty, Trait Anxiety, Threat extinction, Individual Differences, Psychophysiology, Meta-analysis, Multiverse-type analysis

Introduction

In affective neuroscience research there has been a rising interest in examining how individual differences in personality or environmental experiences map on to psychophysiological responses (for examples, see Lonsdorf & Merz, 2017; Morriss et al., 2023; Saarinen et al., 2021; Sigrist et al., 2021). A key advantage of this research is that it may reveal the extent to which different types of individual variation modulate neurobehavioral systems that support affect and arousal (Davidson, 2003; Hariri, 2009). Importantly, the findings from such research can be applied or integrated into existing frameworks of human functioning (e.g. Research Domain Criteria: Insel, 2014) and transdiagnostic models of psychopathology (e.g. The Hierarchical Taxonomy of Psychopathology: Kotov et al., 2017), with the aim to accelerate translation of basic science discoveries to cost-effective and individually tailored therapeutic intervention.

While this continued focus on individual differences in affective neuroscience research is promising, methodological considerations for how to examine individual differences have been somewhat lacking and slower to emerge in some sub fields (Elliott et al., 2021; Yarkoni, 2015), compared to others (e.g. attachment: Gardener et al., 1986; Fraley & Spieker, 2003). Recently, in affective neuroscience research there has been a push to understand how individual differences can be best captured using psychophysiological measures (e.g, see Hajcak et al., 2017). However, there is still a dearth of concrete examples on how different quantification choices for individual differences impact observed relationships between commonly used read-outs (e.g. self-report and psychophysiological responses).

Currently, there are two dominant approaches for quantifying individual differences for statistical analysis (Maxwell & Delaney, 1993): (1) categorical, which involves splitting a set of values into different groups based on the median or one standard deviation above or below the mean, and (2) continuous, which consists of using an entire set of values along a continuum. Categorical and continuous approaches for quantifying individual difference data have several different advantages and disadvantages (for review see, DeCoster et al., 2011). On the one hand, categorical approaches can be useful for probing individual difference data that have skewed distributions and/or extreme scores (e.g. commonly seen for some mental health symptom dimensions), as well as nonlinear relations which are quadratic (e.g. often observed in developmental research) (DeCoster et al., 2011). Although, artificially creating categories will in most cases reduce statistical power (Cohen, 1983), particularly for individual difference predictor variables that are normally distributed (e.g. in a median split, some individuals may be only a few points away from fitting one category over another), as well as increase the chance of false statistical significance when there are multiple predictors (MacCallum et al., 2002; Maxwell & Delaney, 1993). On the other hand, continuous approaches allow for the inclusion of all the available data points and can be applied to most types of models (e.g. linear, nonlinear), which increases statistical power. A disadvantage of the continuous approach for individual difference data is that more participants need to be recruited to achieve an even spread of scores across a given metric.

As far as we are aware there has been no systematic comparison of how categorical and continuous quantification of individual difference data impacts the

observed relationships between different units of analysis (e.g. self-report > psychophysiological responses). There may be benefits to conducting a multiverse-type analysis (Del Giudice & Gangestad, 2021; Steegen et al., 2016) based on different quantification choices for individual difference data. For example, applying both categorical and continuous quantification of individual differences to any given data set may provide useful information about: (1) how to further characterise (e.g. dichotomously; dimensionally) the individual difference, and (2) how to optimally design experiments and prepare analysis pipelines to capture and examine the individual differences across different readouts (e.g. self-report, psychophysiology). Furthermore, investigating the effect of different quantification choices of individual difference data on statistical outcomes (e.g. significance, effect size) will allow for a better assessment of the reliability and reproducibility of the individual difference of interest.

In the current study, we conducted a two-stage aggregated meta-analysis (for discussion see, Boedhoe et al., 2018) on an existing data set (k = 18, n = 1006, Morriss et al., 2021) and compared how the different quantification choices for individual difference variables impacted statistical outcomes such as significance and effect size. The original Morriss et al. (2021) two-stage aggregated meta-analysis examined the relationships between self-reported Intolerance of Uncertainty (IU: the tendency to find uncertainty aversive, Carleton et al., 2007; Freeston et al., 1994) and differential skin conductance response to learned cues during threat extinction training (Lonsdorf et al., 2017). In line with prior research,the total IU scale (27-item, 12-item) and its subscales (inhibitory, prospective) were quantified as continuous variables (Carleton et al., 2007; Freeston et al., 1994). The differential skin conductance response to learned cues

during threat extinction training was indexed by four commonly used metrics (whole phase, early, late, and double-difference [early trials – late trials]) (Fullana et al., 2018; Morriss et al., 2018). To assess the specificity of IU over broader negative affective traits (for discussion see, Morriss, 2023) additional analyses included controlling for self-reported trait anxiety. The meta-analysis revealed that higher IU, across different scales and subscales, was associated with greater skin conductance response to learned threat versus safety cues during the whole and late parts of the threat extinction training phase. Moreover, these IU-related effects remained when controlling for trait anxiety. To interrogate these IU-related effects during threat extinction training further, here we assessed whether these findings replicate when applying different quantification choices for individual differences in IU: categorical via median split; categorical via extremes – one standard deviation above/below; and continuous. This planned analysis may be considered a multiverse-type-analysis as it consists of comparing how different data quantification choices (e.g. categorical versus continuous data) and associated statistical model (e.g. correlations versus ANOVAs) pipelines impact the outcome. The aims of this study run in parallel to a growing literature of multiverse analyses to improve scientific rigor in psychophysiological research generally (Clayson, 2024), and in particular threat conditioning research (Kuhn et al., 2022; Lonsdorf et al., 2019, 2022; Ney et al., 2020, 2022).

Methods

An updated two-stage aggregated meta-analysis was conducted on a pre-existing data set comprising of 18 experiments (n = 1006; see Table 1 and 2) that examined IU and

skin conductance responses during threat extinction training (Goldfarb et al., 2021; Kanen et al., 2021; Lucas et al., 2018; Morriss, 2019; Morriss et al., 2015, 2016, 2020; Morriss & van Reekum, 2019; Sjouwerman et al., 2016, 2020, unpublished data 2020; Steinman et al., 2022; Thompson et al., 2018; de Voogd et al., 2020; Wake et al., 2020, 2021). At the first stage, raw individual level self-reported intolerance of uncertainty/trait anxiety and skin conductance response data from the threat extinction training phase were used to generate summary statistics (i.e. effect sizes) per experiment. At the second stage the summary statistics were synthesised across experiments using fixed-effect meta-analysis models. Please refer to the original two-stage aggregated meta-analysis by Morriss et al. (2021) for details relating to data search and inclusion criteria, data quality checks, and data collation. The experimental parameters and participant characteristics of the experiments are outlined in Table 1 and 2 respectively.

The protocol outlined here was not preregistered. The relevant files from this meta-analysis (i.e., master data file and meta-analysis output from RStudio [RStudio, Inc., Boston, MA]) are located on the Open Science Framework through the following link: https://osf.io/us4qv/

[Insert Table 1 and Table 2 here]

Data reduction

***Intolerance of Uncertainty*:** Scores from four separate Intolerance of Uncertainty Scales (IUS) were generated (IU-27, IU-12, I-IU, and P-IU). The IU-27 consists of 27 items rated on a 5-point Likert scale (Freeston et al., 1994). The IU-12 is generated based on 12 items from the IUS-27 (Carleton et al., 2007). Two experiments collected

the IU-12 and therefore are not included in the analysis of the IU-27 (Lucas et al., 2018; Thompson et al., 2018). The inhibitory IU (I-IU) and Prospective IU (P-IU) subscales measure separate components of IU and are generated from either the IU-27 or the IU-12. Where two or more items were missing for the IUS, values were interpolated based on the average item score (n = 14).

Continuous Measure of IU Scores: For each experiment, the original IU scales and subscales comprised of continuous data and therefore no additional data reduction steps were required.

Median Split of IU Scores: For each experiment, a median was computed for each IUS measure (IU-27, IU-12, I-IU, and P-IU) and then a median split was conducted by dividing the data into two groups (low IU and high IU) based on the median score for each IUS measure (see Fig 1A).

Extreme Values of IU Scores: Within each experiment, the mean and standard deviation was computed for each IUS measure (IU-27, IU-12, I-IU, and P-IU). Subsequently, extreme values for each measure were identified, defined as those deviating from the mean by more than one standard deviation (see Fig 1B). For each experiment, extreme values were categorised as follows: Low IU, representing scores one standard deviation below the mean, and High IU, signifying scores one standard deviation above the mean.

[Insert Fig 1 here]

Trait Anxiety: Of the 18 studies, 15 measured trait anxiety using the State-Trait Anxiety Inventory-Trait (STAI-T: Spielberger et al., 1971) or the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA: Ree et al., 2008). The STAI-T consists of 20 items rated on a 4-point Likert scale, and the STICSA consists of 21 items rated on a 4-point Likert scale.

Skin Conductance Response**:** As in the original paper by Morriss et al. (2021), four separate skin conductance response difference score metrics were computed for each experiment: whole phase extinction [(CS+) − (CS−)], early extinction [(first 6–10 CS+ trials) − (first 6–10 CS− trials)], late extinction [(last 6–10 CS+ trials) − (last 6–10 CS− trials)], and double-difference extinction score $[(CS+ - CS-)_{early} - (CS+ - CS-)_{late}]$. For 4 experiments, only the early extinction training metric was analyzed (Kanen et al., 2021; Sjouwerman et al., 2016, 2020, unpublished data 2020) because these experiments had too few extinction learning trials to compute the other SCR difference score metrics.

Analyses

Correlation and partial correlation analyses were performed in SPSS 19 (IBM Corp.) to generate effect sizes for each experiment based on IU as a continuous variable. The correlations included IU (IU-27, IU-12, I-IU, and P-IU) as a continuous independent variable and the skin conductance difference scores (whole phase, early, late, and double-difference) as a continuous dependent variable. Additionally, partial correlations were conducted on IU (IU-27, IU-12, I-IU, and P-IU) and skin conductance difference

scores (whole phase, early, late, and double-difference), while controlling for trait anxiety.

ANOVAs and ANCOVAs were conducted in RStudio (RStudio, Inc., Boston, MA) to generate effect sizes for each experiment based on IU as a categorical variable (median split or extremes based on one standard deviation above and below the mean). The ANOVAs included IU (IU-27, IU-12, I-IU, and P-IU) as a categorical independent variable and the skin conductance difference scores (whole phase, early, late, and double-difference) as a continuous dependent variable. Furthermore, ANCOVAs were conducted on IU (IU-27, IU-12, I-IU, and P-IU) and skin conductance difference scores (whole phase, early, late, and double-difference), while controlling for trait anxiety.

The *r* values from the correlations and partial correlations using IU as a continuous measure were converted into Hedges' *g* effect size values. The *F* values from ANOVA and ANCOVA analyses using IU as a categorical (i.e., median split and extreme values) measure were also converted into Hedges' *g* effect sizes. Fixed-effect meta-analyses were carried out in RStudio (RStudio, Inc., Boston, MA) on effect sizes across the 18 experiments separately for continuous, median split and extreme measures of IU to generate a pooled effect size for every IU scale/subscale (IU-27, IU-12, I-IU and, P-IU) and difference score (early, late, whole phase, and double-difference). Meta-analyses were repeated for effect sizes calculated when controlling for measures of trait anxiety (STAI and STICSA). Together the different types of analysis resulted in 96 independent effect sizes per experiment (see Table 3).

[Insert Table 3 here]

Benjamini-Hochberg corrections (Benjamini & Hochberg, 1995) were applied for continuous (corrected value, $p < .025$) and categorical (median split corrected value, $p < .046$; extreme values corrected value, $p < .043$) measures of IU. Benjamini-Hochberg corrections were also applied for meta-analyses controlling for trait anxiety (corrected values: continuous, $p < .018$; median split, $p < .040$; extreme value, $p < .043$, measures of IU).

Results

Sections of this text related to the relationship between continuous measures of IU and SCR during threat extinction have been reported in Morriss et al. (2021). For moderator analyses and assessment of publication bias, please refer to Morriss et al. (2021).

For visualisation of the results from the meta-analyses based on the individual difference quantification method see Fig 2 and Supplementary Fig 1 for the effect sizes, and Table 4 for the percentage of significant effects.

[Insert Fig 2 and Table 4 here]

Continuous Measure of IU Scores*:* For IU as a continuous measure, all the self-reported variants of the IUS (IU-27, IU-12, I-IU, and P-IU) were significantly associated with SCR difference scores during late extinction training and across the entire extinction phase (corrected *p*s < .025) (see Fig 2A and Supplementary Table 1). Only the IU-27 (not IU-12, I-IU, or P-IU) was significantly associated with SCR double-

difference scores during extinction training (corrected $p < .025$). None of the self-reported variants of the IUS (IU-27, IU-12, I-IU, and P-IU) were significantly associated with SCR difference scores during early extinction training (*p*s < .3) (See Supplementary Table 1). These analyses yielded small to medium effect sizes (Hedges' *g* 0.20–0.35) with a variable range of heterogeneity (I^2 0–45.6%) depending on the IUS measure (see Fig 2A and Supplementary Table 1).

When controlling for trait anxiety, continuous measures of IU-12, P-IU, and I-IU (but not IU-27) were significantly associated with SCR difference scores during late extinction training (corrected $p < .018$) (See Supplementary Table 4 and Supplementary Fig 1A). Moreover, when controlling for trait anxiety, IU-27, IU-12, and P-IU (but not I-IU) were significantly associated with SCR difference scores across the entire extinction phase (corrected $p < .018$) (See Supplementary Table 4 and Supplementary Fig 1A). Furthermore, when controlling for trait anxiety, none of the self-reported variants of the IUS (IU-27, IU-12, I-IU, and P-IU) were significantly associated with SCR double-difference scores (corrected *p*s between .029 and .06) or SCR difference scores during early extinction training (corrected *p*s > .3) (see Supplementary Table 4 and Supplementary Fig 1A). Again, these analyses produced small to medium effect sizes (Hedges' *g* 0.21–0.31) with a variable range of heterogeneity (I^2 0–52.1%) depending on the IUS measure (see Supplementary Table 4).

Median Split of IU Scores: For IU based on a median split, all the self-reported variants of the IUS (IU-27, IU-12, I-IU, and P-IU) were significantly associated with SCR difference scores during early extinction training, late extinction training and across the

whole extinction phase (corrected *p*s < .046) (see Fig 2B and Supplementary Table 2). IU-27, IU-12 and I-IU were significantly associated with SCR double-difference scores (corrected *p*s < .046), while P-IU was not. The significant meta-analytic effect sizes for relationships between self-reported IU and SCR difference scores during the extinction phase were small to medium (Hedges' *g* 0.26 – 0.47) (see Fig 2B and Supplementary Table 2) and yielded extremely low heterogeneity across studies (I^2 0.0%) (see Supplementary Table 2).

When controlling for trait anxiety, all of the self-reported variants of IU as a median split (IU-27, IU-12, I-IU, and P-IU) were significantly associated with SCR difference scores during early extinction training and across the whole of the extinction phase (corrected *p*s < .040) (see Supplementary Table 5 and Supplementary Fig 1B). IU-27 and IU-12 were significantly associated with SCR difference scores during late extinction training (corrected *p*s < .043), while I-IU and P-IU were not (see Supplementary Table 5 and Supplementary Fig 1B). Further, IU-12 and I-IU were significantly associated with SCR double-difference scores (corrected *p*s < .040), while IU-27 and P-IU were not (see Supplementary Table 5 and Supplementary Fig 1B). The meta-analytic effect sizes for significant relationships between IU based on a median split and SCR difference scores were small to medium (Hedges' *g* 0.21-0.46) and showed extremely low heterogeneity (I^2 0.0%) across studies (see Supplementary Table 5)

Extreme Values of IU Scores: When grouping participants into high vs low IU groups based on extreme values of IU, all the self-reported variants of the IUS (IU-27, IU-12, I-

IU, and P-IU) were significantly associated with SCR difference scores during early and late extinction training and with double difference scores across the extinction phase (corrected *p*s < .043) (see Fig 2C and Supplementary Table 3). IU-27 and P-IU were significantly associated with SCR difference scores across the whole extinction phase (corrected *p*s < .043), while IU-12 and I-IU were not (see Fig 2C and Supplementary Table 3). The significant meta-analytic effect sizes for relationships between IU based on scores one standard deviation below and above the mean and SCR difference scores during the extinction phase were medium to large (Hedges' *g* 0.43 – 0.86) and heterogeneity was variable across studies (I^2 0.0% - 79.6%) (see Supplementary Table 3).

When controlling for trait anxiety, all of the self-reported variants of the IUS as extreme scores (IU-27, IU-12, I-IU, and P-IU) were significantly associated with SCR difference scores during early extinction training, late extinction training and with double difference scores across the extinction phase (corrected *p*s < .043) (see Supplementary Table 6 and Supplementary Fig 1C). IU -27 and P-IU were significantly associated with SCR difference scores across the whole extinction phase (corrected *p*s < .043), while IU-12 and I-IU were not (see Supplementary Table 6 and Supplementary Fig 1C). The significant meta-analytic effect sizes for relationships between IU based on scores one standard deviation below and above the mean and SCR difference scores during the extinction phase, while controlling for trait anxiety, were medium to large (Hedges' *g* 0.43 – 0.99) and heterogeneity was variable across studies (I^2 0.0% - 84.6%) (see Supplementary Table 6 and Supplementary Fig 1C).

[Insert Fig 3 here]

Discussion

The aim of this study was to provide a concrete example of how quantification choices for individual differences impact the observed relationships between different units of analysis (e.g. self-report > psychophysiological responses), in terms of statistical outcomes such as significance and effect size. Here, we conducted an updated two-stage aggregated meta-analysis of self-reported IU and skin conductance responses during threat extinction (k = 18, n = 1006; Morriss et al., 2021) using different quantification choices for individual differences in self-reported IU (categorical via median split, categorical via extreme values – one standard deviation above/below, and continuous). The choice of quantification techniques for individual differences in IU yielded similar and varied meta-analytic results regarding statistical significance and effect sizes.

The different quantification techniques produced some consistent IU-related effects. For instance, across all three individual difference quantification techniques, higher IU, regardless of the scale or subscale used,was significantly associated with larger skin conductance responding to the learned threat vs. safe cues during late extinction training (see Fig 3). Furthermore, the majority of the IU-related effects during late threat extinction training held when controlling for trait anxiety. In addition, a similar pattern emerged for IU across the whole phase of threat extinction training. These findings are in line with prior research demonstrating how higher IU, over other broader negative affective traits, specifically disrupts threat extinction learning (for review see, Morriss et al., 2021).

However, the different quantification techniques also produced varied results. For example, when the IU data were quantified categorically (median split, extremes) but not continuously, higher IU across the different scales and subscales was significantly associated with larger skin conductance responding to the learned threat vs. safety cue during early extinction training. Furthermore, when the IU data were quantified categorically (median split, extreme values) versus continuously, there were more statistically significant effects, and the effect sizes were larger (see Fig 2 and Table 4). Such findings suggest that on the one hand quantifying individual differences categorically may lead to a greater number of type one errors. On the other hand, quantifying individual differences continuously may be the most conservative approach statistically and may be less prone to type two errors. Alternatively, it is possible that the categorical quantification of individual differences in IU produced genuinely unique results. The categorisation of IU data based on extremes of one standard deviation above and below the mean may capture subclinical populations that are more homogenous. Thus, the differences between low and high IU groups may be more likely to be larger and consistent, and hence larger effect sizes occur. In particular, this may explain why the categorisation of IU data based on standard deviation above or below the mean, compared to the other two quantification techniques, also produced the least heterogeneity between the experiments in the meta-analysis.

Taken together, these results suggest that despite different quantification methods for individual differences, self-reported IU broadly, including the total scale and subscales, reliably captures differences in skin conductance responding during threat extinction training. These findings have clear implications for how individual differences

in IU are integrated into transdiagnostic models of the etiology and treatment of anxiety and stress-related conditions (e.g. exposure-based therapies) (for discussion, see Morriss et al., 2021). However, to further understand the translational relevance of IU to threat conditioning mechanisms, future meta-analyses and multiverse-type analyses should aim to investigate whether self-reported IU can be reliably mapped to other psychophysiological metrics (e.g. startle) during threat extinction training. Examining the extent to which IU-related effects during threat extinction training and other threat conditioning phases vary in relation to other individual differences (e.g. personality characteristics, life experiences, developmental windows, mental health conditions) and experimental parameters (e.g. reinforcement rate, types of conditioned/uncondition stimuli) will also be beneficial.

This multiverse-type analysis for the quantification of individual differences sits alongside a recent wave of other multiverse analysis efforts to optimise scientific rigor in psychophysiology generally (Clayson, 2024), and within threat conditioning research (Kuhn et al., 2022; Lonsdorf et al., 2019, 2022; Ney et al., 2020, 2022). As far as we are aware, this is one of the first multiverse-type analysis to examine individual difference quantification techniques and how this influences the relationships between different units of analysis that are thought to capture the same individual difference construct. Overall, in line with previous methodological work (MacCallum et al., 2002; Maxwell & Delaney, 1993), these findings demonstrate that the quantification choices for individual differences impact statistical outcomes such as significance and effect sizes.Furthermore, the current study highlights how conducting a multiverse-type analysis for the different quantification methods of individual differences may help

researchers understand the construct of interest (e.g. characterisation) and how to accurately measure it (e.g. effect sizes for power analyses), as well as examine the stability and reliability of individual difference-based effects and correspondence between various units of analysis (e.g. self-report, psychophysiology). Interrogating individual difference data using this multiverse-type analysis on larger scales will ultimately improve the precision of measuring individual difference constructs, which is crucial for the development of existing frameworks of human functioning (e.g. Research Domain Criteria: Insel, 2014) and transdiagnostic models of psychopathology (e.g. The Hierarchical Taxonomy of Psychopathology: Kotov et al., 2017), to accelerate translation of basic science discoveries to real-world concerns (e.g. clinical practice).

The study did have a few shortcomings. Firstly, we only examined the relationship between two units of analysis, namely self-report and skin conductance response. Future research may wish to expand on this by examining how quantification choices for individual differences impact other combinations of read-outs, which may have their own unique methodological advantages and disadvantages. Secondly, the dataset included in this study comprised of two variables (e.g. self-reported intolerance of uncertainty and skin conductance difference scores) that are often normally distributed, either in their raw form or due to transformation (e.g. square root, z-score). Such patterns of results may not be observed in instances where the variables of interest have non-normal distributions (e.g. skewed, flat, inverted). More research is required to understand how quantification of individual differences impact statistical significance and effect sizes when the variables of interest are non-normally distributed. Thirdly, the study only examined individual difference data in relation to one

experimental paradigm (e.g. threat conditioning experiments) and did not investigate additional experimental parameters as moderators (e.g. reinforcement rate, conditioned stimulus type).Examining these aspects will elucidate whether IU plays a broader role in affective processing more generally or whether IU influences affective processing / threat conditioning processes only under specific circumstances. Fourthly, the study was limited to the usage of cross-sectional data from primarily community/student samples in English-speaking countries. Therefore, further affective neuroscience research using this type of multiverse analysis for the quantification of individual differences is required to assess the specificity, generalisability, and reproducibility of individual difference-based effects in samples with greater demographic diversity (e.g. age, ethnicity, nationality) and samples meeting criteria for clinical presentation of anxiety-related conditions

In conclusion, this study provides a concrete example of how quantification choices for individual differences impact the observed relationships between different units of analysis (e.g. self-report > psychophysiological responses). Future research should compare different quantification choices for individual differences across various units of analysis, in order to advance our understanding of individual difference constructs and measurement, and to realise the benefit of individual differences and psychophysiology research to real-world applications.

References

Boedhoe, P. S., Heymans, M. W., Schmaal, L., Abe, Y., Alonso, P., Ameis, S. H., ... & Twisk, J. W. (2019). An empirical comparison of meta-and mega-analysis with data from the ENIGMA obsessive-compulsive disorder working group. *Frontiers in Neuroinformatics*, 12, 102.

Carleton, R. N., Norton, M. P. J., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*(1), 105-117.

Clayson, P. E. (2024). Beyond single paradigms, pipelines, and outcomes: Embracing multiverse analyses in psychophysiology*. International Journal of Psychophysiology*, 112311.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*(3), 249-253.

Davidson, R. J. (2003). Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology*, *40*(5), 655-665.

DeCoster, J., Gallucci, M., & Iselin, A. M. R. (2011). Best practices for using median splits, artificial categorization, and their continuous alternatives. *Journal of Experimental Psychopathology*, *2*(2), 197-209.

Elliott, M. L., Knodt, A. R., & Hariri, A. R. (2021). Striving toward translation: strategies for reliable fMRI measurement. *Trends in Cognitive Sciences*, *25*(9), 776-787.

Fraley, R. C., & Spieker, S. J. (2003). What are the differences between dimensional and categorical models of individual differences in attachment? Reply to Cassidy

(2003), Cummings (2003), Sroufe (2003), and Waters and Beauchaine (2003). *Developmental Psychology*, *39*(3), 423–429.

Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why do people worry?. *Personality and Individual Differences*, *17*(6), 791-802.

Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., ... & Harrison, B. J. (2018). Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. *Neuroscience & Biobehavioral Reviews*, 88, 16-25.

Gardner, W., Lamb, M. E., Thompson, R. A., & Sagi, A. (1986). On individual differences in strange situation behavior: Categorical and continuous measurement systems in a cross-cultural data set. *Infant Behavior and Development*, *9*(3), 355-375.

Goldfarb, E. V., Blow, T., Dunsmoor, J. E., & Phelps, E. A. (2021). Elemental and configural threat learning bias extinction generalization. *Neurobiology of Learning and Memory*, *180*, 107405.

Hajcak, G., Meyer, A., & Kotov, R. (2017). Psychometrics and the neuroscience of individual differences: Internal consistency limits between-subjects effects. *Journal of Abnormal Psychology*, *126*(6), 823.

Hariri, A. R. (2009). The neurobiology of individual differences in complex behavioral traits. *Annual Review of Neuroscience*, *32*, 225-247.

Insel, T. R. (2014). The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *American Journal of Psychiatry*, *171*(4), 395-397.

Kanen, J. W., Arntz, F. E., Yellowlees, R., Christmas, D. M., Price, A., Apergis-Schoute, A. M., ... & Robbins, T. W. (2021). Effect of tryptophan depletion on conditioned threat memory expression: role of intolerance of uncertainty. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(5), 590-598.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., . . . Clark, L. A. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454.

Kuhn, M., Gerlicher, A. M., & Lonsdorf, T. B. (2022). Navigating the manyverse of skin conductance response quantification approaches-A direct comparison of trough-to-peak, baseline correction, and model-based approaches in Ledalab and PsPM. *Psychophysiology*, 59(9), e14058.

Lonsdorf, T. B., Gerlicher, A., Klingelhöfer-Jens, M., & Krypotos, A. M. (2022). Multiverse analyses in fear conditioning research. *Behaviour Research and Therapy*, 153, 104072.

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., ... & Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, *77*, 247-285.

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans-Biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience & Biobehavioral Reviews*, *80*, 703-728.

Lonsdorf, T. B., Merz, C. J., & Fullana, M. A. (2019). Fear extinction retention: Is it what we think it is? *Biological Psychiatry*, 85(12), 1074-1082.

Lucas, K., Luck, C. C., & Lipp, O. V. (2018). Novelty-facilitated extinction and the reinstatement of conditional human fear. *Behaviour Research and Therapy*, *109*, 68-74.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, *113*(1), 181.

Morriss, J. (2019). What do I do now? Intolerance of uncertainty is associated with discrete patterns of anticipatory physiological responding to different contexts. *Psychophysiology*, *56*(9), e13396.

Morriss, J. (2023). Let's get specific about intolerance of uncertainty and emotion regulation. *Personality and Individual Differences*, 214, 112336.

Morriss, J., Abend, R., Zika, O., Bradford, D. E., & Mertens, G. (2023). Neural and psychophysiological markers of intolerance of uncertainty. *International Journal of Psychophysiology*, *184*, 94-99.

Morriss, J., Christakou, A., & Van Reekum, C. M. (2015). Intolerance of uncertainty predicts fear extinction in amygdala-ventromedial prefrontal cortical circuitry. *Biology of Mood & Anxiety Disorders*, *5*, 1-13.

Morriss, J., Christakou, A., & Van Reekum, C. M. (2016). Nothing is safe: Intolerance of uncertainty is associated with compromised fear extinction learning. *Biological Psychology*, *121*, 187-193.

Morriss, J., Hoare, S., & van Reekum, C. M. (2018). It's time: A commentary on fear extinction in the human brain using fMRI. *Neuroscience & Biobehavioral Reviews*, 94, 321-322.

Morriss, J., & van Reekum, C. M. (2019). I feel safe when i know: Contingency instruction promotes threat extinction in high intolerance of uncertainty individuals. *Behaviour Research and Therapy*, *116*, 111-118.

Morriss, J., Wake, S., Elizabeth, C., & Van Reekum, C. M. (2021). I doubt it is safe: A meta-analysis of self-reported intolerance of uncertainty and threat extinction training. *Biological Psychiatry Global Open Science*, *1*(3), 171-179.

Morriss, J., Wake, S., Lindner, M., McSorley, E., & Dodd, H. (2020). How many times do I need to see to believe? The impact of intolerance of uncertainty and exposure experience on safety-learning and retention in young adults. *International Journal of Psychophysiology*, *153*, 8-17.

Morriss, J., Zuj, D. V., & Mertens, G. (2021). The role of intolerance of uncertainty in classical threat conditioning: Recent developments and directions for future research. *International Journal of Psychophysiology*, 166, 116-126.

Ney, L. J., Laing, P. A. F., Steward, T., Zuj, D. V., Dymond, S., & Felmingham, K. (2020). Inconsistent analytic strategies reduce robustness in fear extinction via skin conductance response. *Psychophysiology*, 57(11), e13650

Ney, L. J., Laing, P. A. F., Steward, T., Zuj, D. V., Dymond, S., Harrison, B., Graham, B., & Felmingham, K. L. (2022). Methodological implications of sample size and extinction gradient on the robustness of fear conditioning across different analytic strategies. *PLOS ONE*, 17(5), e0268814

Ree, M. J., French, D., MacLeod, C., & Locke, V. (2008). Distinguishing cognitive and somatic dimensions of state and trait anxiety: Development and validation of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). *Behavioural and Cognitive Psychotherapy*, *36*(3), 313-332.

Saarinen, A., Keltikangas-Järvinen, L., Jääskeläinen, E., Huhtaniska, S., Pudas, J., Tovar-Perdomo, S., . . . Lieslehto, J. (2021). Early adversity and emotion processing from faces: a meta-analysis on behavioral and neurophysiological responses. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(7), 692-705.

Sigrist, C., Mürner-Lavanchy, I., Peschel, S. K., Schmidt, S. J., Kaess, M., & Koenig, J. (2021). Early life maltreatment and resting-state heart rate variability: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *120*, 307-334.

Sjouwerman, R., Niehaus, J., Kuhn, M., & Lonsdorf, T. B. (2016). Don't startle me—Interference of startle probe presentations and intermittent ratings with fear acquisition. *Psychophysiology*, *53(*12), 1889-1899.

Sjouwerman, R., Scharfenort, R., & Lonsdorf, T. B. (2020). Individual differences in fear acquisition: multivariate analyses of different emotional negativity scales, physiological responding, subjective measures, and neural activation. *Scientific Reports*, *10(*1), 15283.

Spielberger, C. D., Gonzalez-Reigosa, F., Martinez-Urrutia, A., Natalicio, L. F., & Natalicio, D. S. (1971). The state-trait anxiety inventory. *Interamerican Journal of Psychology*, *5*(3 & 4).

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702-712.

Steinman, S. A., Dunsmoor, J. E., Gazman, Z., Stovezky, Y., Pascucci, O., Pomerenke, J., ... & Simpson, H. B. (2022). A Preliminary Test of Novelty-Facilitated Extinction in Individuals With Pathological Anxiety. *Frontiers in Behavioral Neuroscience*, *16*, 873489.

Thompson, A., McEvoy, P. M., & Lipp, O. V. (2018). Enhancing extinction learning: Occasional presentations of the unconditioned stimulus during extinction eliminate spontaneous recovery, but not necessarily reacquisition of fear. *Behaviour Research and Therapy*, *108*, 29-39.

de Voogd, L. D., & Phelps, E. A. (2020). A cognitively demanding working-memory intervention enhances extinction. *Scientific Reports*, *10*(1), 7020.

Wake, S., Morriss, J., Johnstone, T., Van Reekum, C. M., & Dodd, H. (2021). Intolerance of uncertainty, and not social anxiety, is associated with compromised extinction of social threat. *Behaviour Research and Therapy*, *139*, 103818.

Wake, S., van Reekum, C. M., Dodd, H., & Morriss, J. (2020). The impact of intolerance of uncertainty and cognitive behavioural instructions on safety learning. *Cognitive Therapy and Research*, *44*, 931-942.

Yarkoni, T. (2015). Neurobiological substrates of personality: A critical overview. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), APA handbook of personality and social psychology, Vol. 4. *Personality processes and individual differences* (pp. 61–83). American Psychological Association.

Declarations of interest: none

Table 1.

Experimental parameters across threat conditioning studies

Study	IU Scale Administered	Trait Anxiety Measure Administered	Reinforcement Rate	Instruction Type	CS Type	US Type	CS Length (ms)	ITI Length (ms)	N Trials Extinction	SCR Scoring Window (ms after trial onset)
Goldfarb et al. (2021)	IU-27	STAI-T	73%	Uninstructed	Tones and coloured squares	Electric shock	6000	8000 – 10000	24 (12+, 12 CS-)	500 - 6000
Kanen et al. (2021)	IU-27	STAI-T	37.5%	Uninstructed	Coloured squares	Electric shock	4000	10000	20 (10 CS+, 10 CS-)	500 - 4500
Lucas et al. (2018)	IU-12	N/A	50%	Uninstructed	Angry male white faces	Electric shock	8000	22000, 24000 or 26000	32 (16 CS+, 16 CS-)	1000 – 4000
Morriss et al. (2015)	IU-27	STAI-T	100%	Uninstructed	Coloured squares	Female scream	1500	3000 – 6450	32 (16 CS+, 16 CS-)	0 – 7000
Morriss et al. (2016)	IU-27	STAI-T	100%	Uninstructed	Coloured squares	Female scream	1500	3000 – 6450	32 (16 CS+, 16 CS-)	0 – 7000
Morriss & van Reekum (2019) Exp 1	IU-27	STAI-T	50%	Uninstructed	Coloured squares	Female scream	4000	6000 – 8800	32 (16 CS+, 16 CS-)	500 – 3500
Morriss & van Reekum (2019) Exp 2	IU-27	STAI-T	50%	Uninstructed	Coloured squares	Female scream	4000	6000 – 8800	32 (16 CS+, 16 CS-)	500 – 3500
Morriss & van Reekum (2019) Exp 3	IU-27	STAI-T	50%	Uninstructed	Coloured squares	Female scream	4000	6000 – 8800	32 (16 CS+, 16 CS-)	500 – 3500
Morriss (2019)	IU-27	STAI-T	50%	Uninstructed	Coloured squares	Female scream	4000	6000 – 8800	32 (16 CS+, 16 CS-)	500 – 3500
	IU-27	STICSA	50%	Uninstructed			4000			

Morriss et al. (2019)					Coloured squares	Female scream		6000 – 8800	32 (16 CS+, 16 CS-)	500 – 3500
Sjouwerman et al. (2016)	IU-27	STAI-T	100%	Uninstructed	Black shapes (i.e. grid or spiral) on a background picture of water	Electric shock	6000	10000 – 13000	18 (9 CS+, CS-)	900 - 4000
Sjouwerman et al. (2020)	IU-27	STAI-T	100%	Uninstructed	Black geometrical symbols on coloured background	Electric shock	6000	10000 - 13000	18 (9 CS+, CS-)	900 - 4000
Sjouwerman & Lonsdorf (Unpublished Data)	IU-27	STAI-T	100%	Uninstructed	Black geometrical symbols on coloured background	Electric shock	6000	11000 - 13000	18 (9 CS+, CS-)	900 - 3500
Steinman et al. (2022)	IU-27	STAI-T	53.33%	Uninstructed	Angry male white faces	Electric shock	6000	12000	40 (20 CS+, 20 CS-)	500 - 6000
Thompson et al. (2018)	IU-12	N/A	100%	Uninstructed	Coloured images of animals (fish and birds)	Electric shock	6000	13000 - 17000	24 (12 CS+, 12 CS-)	1000 - 6000
de Voogd et al. (2020)	IU-27	N/A	37.50%	Uninstructed	Pictures of snakes	Electric shock	6000	18000 – 22000	40 (20 CS+, 20 CS-)	500 - 6500
Wake et al. (2020)	IU-27	STICSA	50%	Uninstructed	Coloured squares	Female scream	4000	6000 – 8800	32 (16 CS+, 16 CS-)	500 – 3500
Wake et al. (2021)	IU-27	STAI-T	50%	Uninstructed	Neutral female white faces	Electric shock and critical verbal statements	4000	6000 – 8800	32 (16 CS+, 16 CS-)	500 – 3500

Intolerance of Uncertainty, IU-12 or IU-27; STAI-T, State-Trait Anxiety Inventory-Trait; STICSA, The State-Trait Inventory for Cognitive and Somatic Anxiety; CS, Conditioned stimulus; US, Unconditioned stimulus; N, Number; ITI, Inter-trial interval; SCR, Skin Conductance Response.

Table 2

Participant characteristics across threat conditioning studies

Study	**Sample Type**	**Sex**	**Ethnicity**	**Age**
Goldfarb et al. (2021)	Community and students	30 F / 18 M	Not recorded	22.25
Kanen et al. (2021)	Non-clinical, community	18 F / 29 M	Data not returned on time	25
Lucas et al. (2018)	Community and students	29 F / 19 M (across exp groups)	15 Caucasian, 7 Asian, 2 Indian	25 (across exp groups)
Morriss et al. (2015)	Community and students	12 F / 10 M	18 White, 2 Asian, 2 Mixed	23.59
Morriss et al. (2016)	Students	32 F / 6 M	Not recorded	18-25 years
Morriss & van Reekum (2019) Exp 1	Community and students	33 F / 27 M (across exp groups)	Not recorded	23.56 (across exp groups)
Morriss & van Reekum (2019) Exp 2	Community and students	57 F / 24 M (across exp groups)	Not recorded	24.65 (across exp groups)
Morriss & van Reekum (2019) Exp 3	Students	86 F / 11 M (across exp groups)	72 White, 13 Asian, 6 Black, 4 Mixed, 2 Middle Eastern (across exp groups)	20.61 (across exp groups)
Morriss (2019)	Community and students	31 F / 14 M	33 White, 5 Asian, 4 Black, 3 Mixed	23
Morriss et al. (2019)	Community and students	86 F / 58 M (across exp groups)	92 White, 29 Asian, 15 not specified, 4 Middle Eastern/Arab, 2 Black, 2 Mixed (across exp groups)	24 (across exp groups)
Sjouwerman et al. (2016)	Community and students	255 F / 101 M	Not recorded	25
Sjouwerman et al. (2020)	Community and students	38 F / 19 M	Not recorded	25

Sjouwerman & Lonsdorf (Unpublished Data)	Community and students	66 F / 22 M	Not recorded	25
Steinman et al. (2022)	Clinically diagnosed with anxiety or obsessive compulsive disorder	15 F / 12 M	17 White, 4 Other, 3 Black, 3 Asian	24.33
Thompson et al. (2018)	Students	15 F / 9 M	15 Caucasian, 7 Asian, 2 Indian	20.37
de Voogd et al. (2020)	Non-clinical, students	61 F/ 41 M (across exp groups)	35 Asian/Asian Americans, 29 Caucasian/White, 20 Black/African American, 8 Mixed, 5 Unknown/not indicated, 3 Hispanic non-white, 1 Hispanic White, 1 Arab (across exp groups)	24.4 (across exp groups)
Wake et al. (2020)	Community and students	67 F / 28 M (across exp groups)	61 White, 14 not specified, 10 Asian, 3 Middle Eastern/Arab, 2 Black, 1 Mixed (across exp groups).	24.4 (across exp groups)
Wake et al. (2021)	Students	84 F	67 White, 14 Asian/Pacific Islander, 7 Black, 3 Mixed, 1 Middle Eastern/Arab	19.66

Table 3.

Analysis steps

Quantification	**Statistical Test**	**IU scales**	**Skin conductance difference scores**	**Control for trait anxiety**	**Total independent effect sizes generated**
Continuous	Correlation	4: IU-27, IU-12, I-IU, and P-IU	4: whole phase, early, late, and double-difference		16
Continuous	Partial Correlation	4: IU-27, IU-12, I-IU, and P-IU	4: whole phase, early, late, and double-difference	x	16
Median Split	ANOVA	4: IU-27, IU-12, I-IU, and P-IU	4: whole phase, early, late, and double-difference		16
Median Split	ANCOVA	4: IU-27, IU-12, I-IU, and P-IU	4: whole phase, early, late, and double-difference	x	16
Extreme Values	ANOVA	4: IU-27, IU-12, I-IU, and P-IU	4: whole phase, early, late, and double-difference		16
Extreme Values	ANCOVA	4: IU-27, IU-12, I-IU, and P-IU	4: whole phase, early, late, and double-difference	x	16

Table 4.

Percentage of significant effects from the meta-analyses based on the individual difference quantification method

	Continuous	Median Split	Extreme Values
IU	56.25%	93.75%	87.50%
IU controlling for trait anxiety	56.25%	81.25%	87.50%

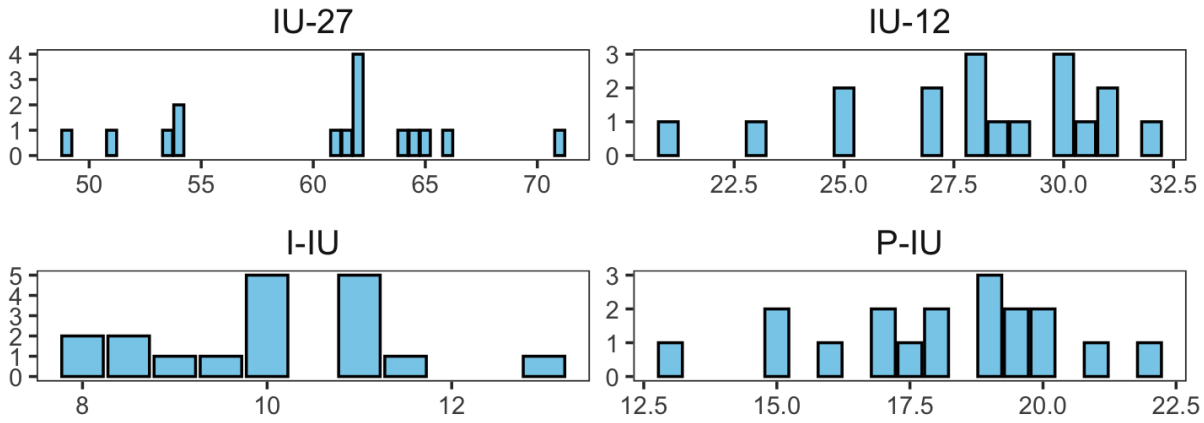
Figure captions

Fig 1. Histograms of the median **(A)** and standard deviations **(B)** for each IUS measure (IUS-27, IU-12, I-IU, P-IU). The X axis for **A** represents the median spit of IU scores and for **B** represents the extreme values of IU scores – one standard deviation above and below the mean. The Y axis represents the frequency of experiments.

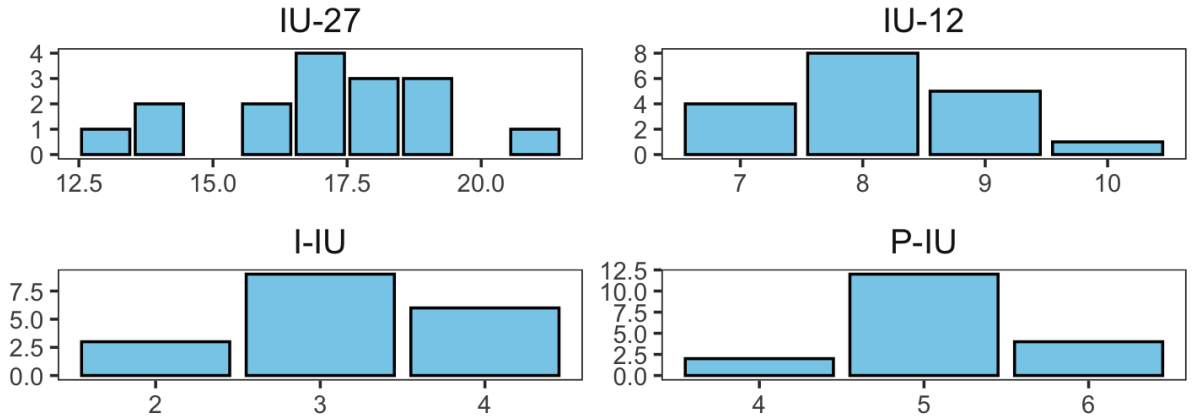
Fig 2. A plot depicting the meta-analytic pooled effect sizes for every IU scale/subscale (IU-27, IU-12, I-IU and, P-IU) and skin conductance response difference scores (CS+ - CS-) during threat extinction (early, late, whole phase, and double-difference) based on the individual difference quantification method for IU (**A** = continuous, **B** = median split, **C** = extreme values – one standard deviation above and below the mean). The values in the cells represent hedges' *g* effect sizes. Empty cells represent non-significant effects.

Fig 3. Forest plots demonstrating effect sizes across studies for the relationships between the 12-item Intolerance of Uncertainty Scale (IU-12) and skin conductance response difference scores (CS+ - CS-) during late extinction training when IU-12 is quantified continuously **(A)**, based on a median split **(B)**, and based on extreme values – one standard deviation above and below the mean **(C)**. CI, confidence interval; SMD, standardised mean difference as hedges' *g* effect sizes.

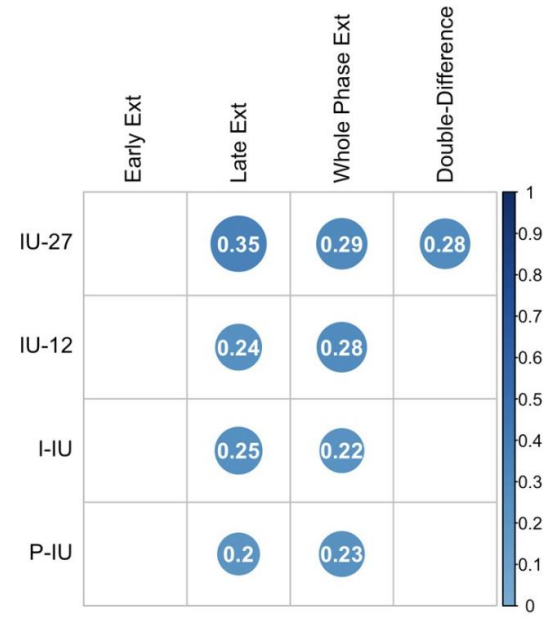
A



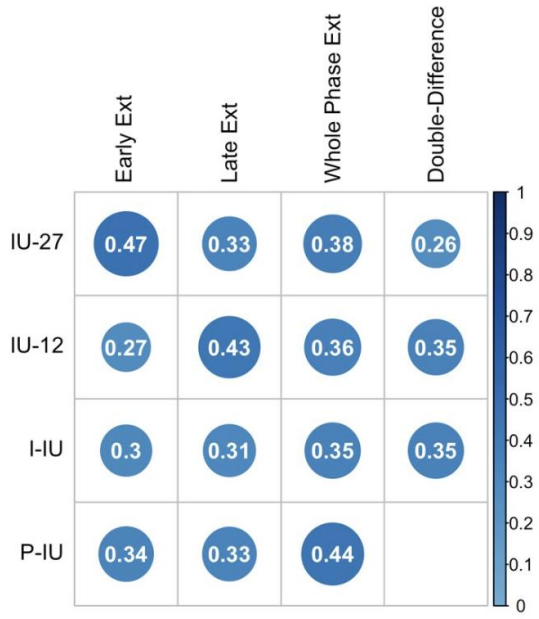
B



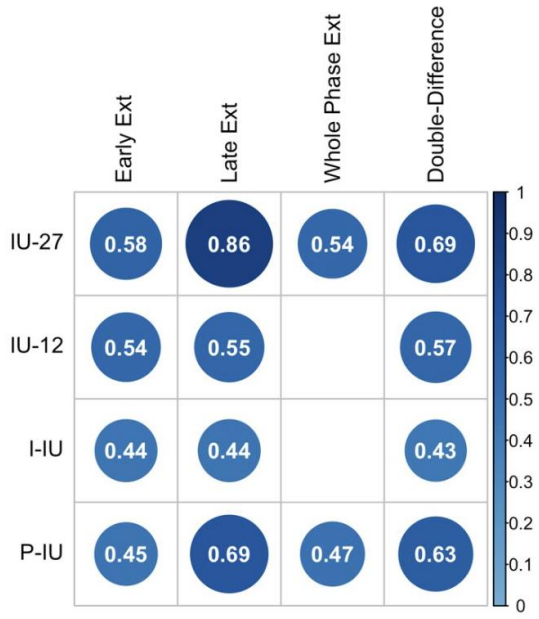
A. Continuously Measured IU



B. Median Split IU



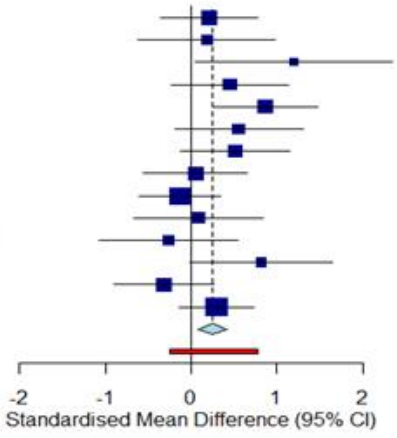
C. Extreme Values IU



A

Source	SMD (95% CI)
Goldfarb et al. (2021)	0.21 [-0.35; 0.77]
Lucas et al. (2018)	0.18 [-0.62; 0.99]
Morriss et al. (2015)	1.20 [0.05; 2.35]
Morriss et al. (2016)	0.45 [-0.23; 1.13]
Morriss (2019)	0.87 [0.26; 1.47]
Morriss & van Reekum (2019), Exp 1	0.56 [-0.18; 1.30]
Morriss & van Reekum (2019), Exp 2	0.51 [-0.12; 1.15]
Morriss & van Reekum (2019), Exp 3	0.05 [-0.56; 0.65]
Morriss et al. (2020)	-0.13 [-0.60; 0.34]
Steinman et al. (2022)	0.09 [-0.67; 0.84]
Thompson et al. (2018)	-0.26 [-1.07; 0.54]
de Voogd et al. (2020)	0.82 [-0.02; 1.65]
Wake et al. (2020)	-0.32 [-0.90; 0.26]
Wake et al. (2021)	0.30 [-0.13; 0.73]
Total	0.24 [0.08; 0.41]
Prediction interval	[-0.26; 0.78]

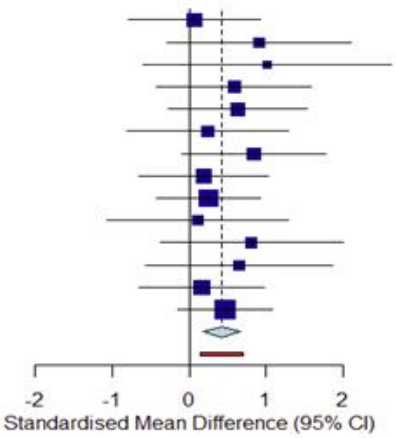
Heterogeneity: $\chi^2_{13} = 18.56$ ($P = .14$), $I^2 = 30\%$



B

Source	SMD (95% CI)
Goldfarb et al. (2021)	0.07 [-0.79; 0.92]
Lucas et al. (2018)	0.91 [-0.29; 2.11]
Morriss et al. (2015)	1.01 [-0.60; 2.62]
Morriss et al. (2016)	0.58 [-0.42; 1.59]
Morriss (2019)	0.63 [-0.27; 1.53]
Morriss & van Reekum (2019), Exp 1	0.24 [-0.81; 1.29]
Morriss & van Reekum (2019), Exp 2	0.84 [-0.10; 1.78]
Morriss & van Reekum (2019), Exp 3	0.19 [-0.65; 1.03]
Morriss et al. (2020)	0.26 [-0.42; 0.93]
Steinman et al. (2022)	0.11 [-1.07; 1.30]
Thompson et al. (2018)	0.81 [-0.37; 2.00]
de Voogd et al. (2020)	0.65 [-0.57; 1.87]
Wake et al. (2020)	0.16 [-0.65; 0.98]
Wake et al. (2021)	0.47 [-0.15; 1.09]
Total	0.43 [0.18; 0.67]
Prediction interval	[0.15; 0.70]

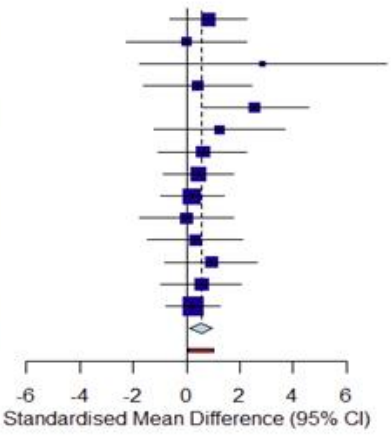
Heterogeneity: $\chi^2_{13} = 4.74$ ($P = .98$), $I^2 = 0\%$



C

Source	SMD (95% CI)
Goldfarb et al. (2021)	0.81 [-0.65; 2.27]
Lucas et al. (2018)	0.00 [-2.27; 2.27]
Morriss et al. (2015)	2.85 [-1.78; 7.49]
Morriss et al. (2016)	0.43 [-1.60; 2.45]
Morriss (2019)	2.56 [0.56; 4.56]
Morriss & van Reekum (2019), Exp 1	1.23 [-1.23; 3.69]
Morriss & van Reekum (2019), Exp 2	0.61 [-1.05; 2.26]
Morriss & van Reekum (2019), Exp 3	0.44 [-0.88; 1.77]
Morriss et al. (2020)	0.20 [-0.99; 1.38]
Steinman et al. (2022)	0.00 [-1.75; 1.75]
Thompson et al. (2018)	0.32 [-1.45; 2.10]
de Voogd et al. (2020)	0.93 [-0.80; 2.65]
Wake et al. (2020)	0.55 [-0.97; 2.08]
Wake et al. (2021)	0.23 [-0.79; 1.24]
Total	0.55 [0.12; 0.98]
Prediction interval	[0.07; 1.03]

Heterogeneity: $\chi^2_{13} = 6.85$ ($P = .91$), $I^2 = 0\%$



Supplementary Table 1.
Pooled effect sizes and heterogeneity[a] using continuous IU scales and difference scores for SCR during extinction.

	g	(95% CI)	*k*	*N*	*p*	I^2
IU-27						
Early Ext	-0.01	(-0.14; 0.11)	16	958	.822	45.6%
Late Ext	0.35	(0.17; 0.53)	12	504	<.001	17.2%
Whole Phase Ext	0.29	(0.11; 0.46)	12	504	.001	1.5%
Double-Difference	-0.28	(-0.45; -0.1)	12	504	.002	31.8%
IU-12						
Early Ext	0.06	(-0.06; 0.19)	18	1006	.341	44.9%
Late Ext	0.24	(0.08; 0.41)	14	552	.005	29.9%
Whole Phase Ext	0.28	(0.11; 0.45)	14	552	.001	1.1%
Double-Difference	-0.12	(-0.28; 0.05)	14	552	.180	38.2%
I-IU						
Early Ext	0.03	(-0.09; 0.16)	18	1006	.621	40.4%
Late Ext	0.25	(0.08; 0.42)	14	552	.004	10.9%
Whole Phase Ext	0.22	(0.05; 0.39)	14	552	.010	0%
Double-Difference	-0.15	(-0.32; 0.02)	14	552	.090	43%
P-IU						
Early Ext	0.07	(-0.06; 0.19)	18	1006	.301	42%
Late Ext	0.20	(0.04; 0.37)	14	552	.017	32.5%
Whole Phase Ext	0.23	(0.06; 0.39)	14	552	.008	10%
Double-Difference	-0.08	(-0.25; 0.09)	14	552	.348	20%

CI, confidence interval; ext, extinction; I-IU, Inhibitory IU; IU, intolerance of uncertainty; P-IU, Prospective IU; SCR, skin conductance response.
[a] Percentage of variability in effect size

Supplementary Table 2.
Pooled effect sizes and heterogeneity[a] using median split IU scales and difference scores for SCR during extinction.

	g	(95% CI)	*k*	*N; grp1, grp2*	*p*	I^2
IU-27						
Early Ext	0.47	0.29, 0.66	16	463, 494	<.001	0.0%
Late Ext	0.33	0.08, 0.59	12	244, 259	.010	0.0%
Whole Phase Ext	0.38	0.13, 0.64	12	244, 259	.003	0.0%
Double-Difference	0.26	0.01, 0.52	12	244, 259	.042	0.0%
IU-12						
Early Ext	0.27	0.09, 0.46	18	473, 532	.003	0.0%
Late Ext	0.43	0.18, 0.67	14	263, 288	.001	0.0%
Whole Phase Ext	0.36	0.12, 0.61	14	263, 288	.004	0.0%
Double-Difference	0.35	0.11, 0.60	14	263, 288	.005	0.0%
I-IU						
Early Ext	0.30	0.12, 0.49	18	470, 535	.001	0.0%
Late Ext	0.31	0.06, 0.56	14	253, 298	.014	0.0%
Whole Phase Ext	0.35	0.10, 0.60	14	253, 298	.007	0.0%
Double-Difference	0.35	0.10, 0.60	14	253, 298	.007	0.0%
P-IU						
Early Ext	0.34	0.16, 0.53	18	461, 544	<.001	0.0%
Late Ext	0.33	0.09, 0.58	14	261, 290	.008	0.0%
Whole Phase Ext	0.44	0.19, 0.68	14	261, 290	.001	0.0%
Double-Difference	0.24	-0.00, 0.49	14	261, 290	.054	0.0%

CI, confidence interval; ext, extinction; I-IU, Inhibitory IU; IU, intolerance of uncertainty; P-IU, Prospective IU; SCR, skin conductance response.
[a] Percentage of variability in effect size

Supplementary Table 3.
Pooled effect sizes and heterogeneity[a] (percentage of variability in the effect size) using extreme values of IU scales (IU-27, IU-12, I-IU, and P-IU) and difference scores (early, late, whole phase and double-difference) for SCR during extinction.

	g	(95% CI)	*k*	*N; grp1, grp2*	*p*	I^2
IU-27						
Early Ext	0.58	0.25, 0.91	16	153, 169	.001	32.6%
Late Ext	0.86	0.40, 1.33	12	83, 88	<.001	63.7%
Whole Phase Ext	0.54	0.09, 0.99	12	83, 88	.018	0.0%
Double-Difference	0.69	0.23, 1.14	12	83, 88	.003	79.6%
IU-12						
Early Ext	0.54	0.22, 0.85	18	162, 171	.001	0.0%
Late Ext	0.55	0.12, 0.98	14	92, 98	.013	0.0%
Whole Phase Ext	0.36	-0.07, 0.78	14	92, 98	.097	0.0%
Double-Difference	0.57	0.13, 1.00	14	92, 98	.010	0.0%
I-IU						
Early Ext	0.44	0.15, 0.72	18	201, 172	.003	0.0%
Late Ext	0.44	0.05, .082	14	110, 99	.027	0.0%
Whole Phase Ext	0.36	-0.02, 0.74	14	110, 99	.065	0.0%
Double-Difference	0.43	0.05, 0.82	14	110, 99	.028	0.0%
P-IU						
Early Ext	0.45	0.14, 0.76	18	167, 172	.005	0.0%
Late Ext	0.69	0.26, 1.11	14	95, 96	.002	0.0%
Whole Phase Ext	0.47	0.05, 0.88	14	95, 96	.029	0.0%
Double-Difference	0.63	0.21, 1.06	14	95, 96	.003	0.0%

CI, confidence interval; ext, extinction; I-IU, Inhibitory IU; IU, intolerance of uncertainty; P-IU, Prospective IU; SCR, skin conductance response.
[a] Percentage of variability in effect size

Supplementary Table 4.
Pooled effect sizes and heterogeneity[a] using continuous IU scales and difference scores for SCR during extinction when controlling for trait anxiety scores.

	g	(95% CI)	*k*	*N*	*p*	I^2
IU-27						
Early Ext	-0.03	(-0.16; 0.1)	15	933	.661	45.3%
Late Ext	0.14	(-0.04; 0.33)	11	479	.120	65.5%
Whole Phase Ext	0.31	(0.13; 0.49)	11	479	<.001	18.8%
Double-Difference	-0.21	(-0.39; -0.02)	11	479	.029	52.1%
IU-12						
Early Ext	0.04	(-0.08; 0.17)	15	933	.498	46.1%
Late Ext	0.28	(0.10; 0.47)	11	479	.002	26.9%
Whole Phase Ext	0.25	(0.07; 0.43)	11	479	.007	14%
Double-Difference	-0.17	(-0.35; 0.01)	11	479	.066	43%
I-IU						
Early Ext	0.004	(-0.12; 0.13)	15	933	.947	41.2%
Late Ext	0.29	(0.11; 0.47)	11	479	.002	18%
Whole Phase Ext	0.18	(0.002; 0.36)	11	479	.047	16.1%
Double-Difference	-0.18	(-0.37; -0.003)	11	479	.046	51.6%
P-IU						
Early Ext	0.06	(-0.07; 0.19)	15	933	.354	32.8%
Late Ext	0.23	(0.05; 0.41)	11	479	.013	22%
Whole Phase Ext	0.24	(0.06; 0.42)	11	479	.009	0%
Double-Difference	-0.13	(-0.31; 0.05)	11	479	.163	19.3%

CI, confidence interval; ext, extinction; I-IU, Inhibitory IU; IU, intolerance of uncertainty; P-IU, Prospective IU; SCR, skin conductance response.

[a] Percentage of variability in effect size

[b] State-Trait Anxiety Inventory-Trait or State-Trait Inventory for Cognitive and Somatic Anxiety.

Supplementary Table 5.
Pooled effect sizes and heterogeneity[a] using median split IU scales and difference scores for SCR during extinction when controlling for trait anxiety scores[b].

	g	(95% CI)	*k*	*N; grp1, grp2*	*p*	I^2
IU-27						
Early Ext	0.21	0.01, 0.41	13	388, 416	.039	0.0%
Late Ext	0.40	0.09, 0.70	9	175, 188	.010	0.0%
Whole Phase Ext	0.46	0.15, 0.76	9	175, 188	.003	0.0%
Double-Difference	0.25	-0.05, 0.55	9	175, 188	.096	0.0%
IU-12						
Early Ext	0.26	0.05, 0.46	13	375, 429	.014	0.0%
Late Ext	0.43	0.12, 0.73	9	171, 192	.006	0.0%
Whole Phase Ext	0.41	0.10, 0.71	9	171, 192	.009	0.0%
Double-Difference	0.37	0.06, 0.67	9	171, 192	.019	0.0%
I-IU						
Early Ext	0.30	0.10, 0.51	13	374, 430	.004	0.0%
Late Ext	0.32	0.01, 0.63	9	163, 200	.046	0.0%
Whole Phase Ext	0.38	0.07, 0.69	9	163, 200	.016	0.0%
Double-Difference	0.38	0.07, 0.69	9	163, 200	.016	0.0%
P-IU						
Early Ext	0.36	0.15, 0.57	13	363, 441	.001	0.0%
Late Ext	0.30	-0.01, 0.60	9	168, 195	.057	0.0%
Whole Phase Ext	0.44	0.13, 0.75	9	168, 195	.005	0.0%
Double-Difference	0.21	-0.10, 0.51	9	168, 195	.181	0.0%

CI, confidence interval; ext, extinction; I-IU, Inhibitory IU; IU, intolerance of uncertainty; P-IU, Prospective IU; SCR, skin conductance response.
[a] Percentage of variability in effect size
[b] State-Trait Anxiety Inventory-Trait or State-Trait Inventory for Cognitive and Somatic Anxiety.

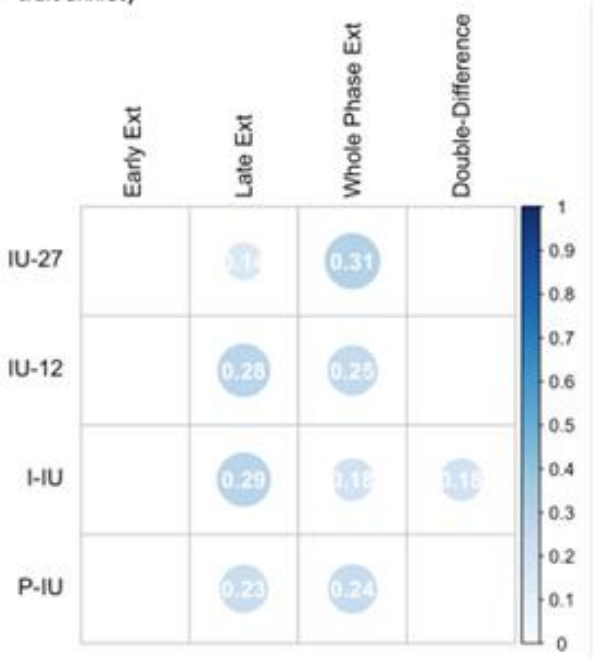
Supplementary Table 6.
Pooled effect sizes and heterogeneity[a] (percentage of variability in the effect size) using extreme values of IU scales (IU-27, IU-12, I-IU, and P-IU) and difference scores (early, late, whole phase and double-difference) for SCR during extinction when controlling for trait anxiety scores (STAI-T or STICSA)[b].

	g	(95% CI)	*k*	*N; grp1, grp2*	*p*	I^2
IU-27						
Early Ext	0.62	0.25, 0.98	13	125, 145	.001	43.3%
Late Ext	0.99	0.44, 1.56	9	58, 67	.001	73.0%
Whole Phase Ext	0.62	0.07, 1.17	9	58, 67	.026	0.0%
Double-Difference	0.69	0.15, 1.24	9	58, 67	.013	84.6%
IU-12						
Early Ext	0.52	0.17, 0.88	13	128, 134	.004	0.0%
Late Ext	0.63	0.09, 1.18	9	60, 65	.022	0.0%
Whole Phase Ext	0.37	-0.15, 0.89	9	60, 65	.168	0.0%
Double-Difference	0.62	0.08, 1.15	9	60, 65	.024	0.0%
I-IU						
Early Ext	0.43	0.12, 0.75	13	161, 137	.007	0.0%
Late Ext	0.54	0.07, 1.01	9	74, 67	.026	0.0%
Whole Phase Ext	0.36	-0.11, 0.83	9	74, 67	.130	0.0%
Double-Difference	0.52	0.05, 0.99	9	74, 67	.032	0.0%
P-IU						
Early Ext	0.50	0.15, 0.85	13	132, 139	.005	0.0%
Late Ext	0.80	0.26, 1.33	9	62, 66	.004	0.0%
Whole Phase Ext	0.60	0.07, 1.12	9	62, 66	.025	0.0%
Double-Difference	0.74	0.21, 1.27	9	62, 66	.007	0.0%

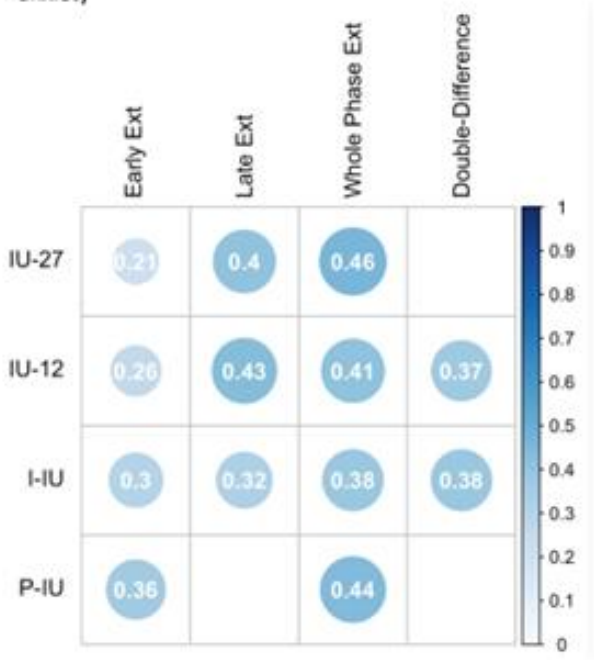
CI, confidence interval; ext, extinction; I-IU, Inhibitory IU; IU, intolerance of uncertainty; P-IU, Prospective IU; SCR, skin conductance response.
[a] Percentage of variability in effect size
[b] State-Trait Anxiety Inventory-Trait or State-Trait Inventory for Cognitive and Somatic Anxiety.

Supplementary Fig 1. A plot depicting the meta-analytic pooled effect sizes for every IU scale/subscale (IU-27, IU-12, I-IU and, P-IU) and skin conductance response difference scores (CS+ - CS-) during threat extinction (early, late, whole phase, and double-difference) whilst controlling for trait anxiety, based on the individual difference quantification method for IU (**A** = continuous, **B** = median split, **C** = extreme values – one standard deviation above and below the mean). The values in the cells represent hedges' *g* effect sizes. Empty cells represent non-significant effects.

A. Continuously Measured IU, after controlling for trait anxiety



B. Median Split IU,, after controlling for trait anxiety



C. Extreme Values IU,, after controlling for trait anxiety