

# Semantic Factor Analysis: Validating Personality Structure Recovery from empirically-weighted Word Embeddings

Oliver Müller

August-Wilhelm Scheer Institut

oliver.mueller@aws-institut.de

## Abstract

The present study introduces Semantic Factor Analysis (SFA), a novel computational approach recovering Big Five personality trait structures from pre-trained adjective word embeddings weighted by empirical participant data. Using Word2Vec embeddings trained on Google-News-300 corpus, semantic relationships of IPIP-50 Big Five inventory adjectives (Goldberg, 1992) were extracted and factor structures computed through weighted vector averaging and K-means clustering. To validate the methodology, SFA was compared against a baseline using unweighted Word2Vec embeddings. In a controlled experiment with  $n=55$  participants completing standard IPIP-50 assessments, HSP-R scale (Pluess et al., 2024) and multimedia impact surveys, empirically-weighted SFA successfully recovered all five personality dimensions with 62.5% average factor purity, substantially outperforming the unweighted baseline (52.0%, 10% relative improvement), while traditional Confirmatory Factor Analysis showed factor collapse and poor model fit. The approach was validated through Latent Class Analysis deriving empirically-based classification thresholds for Big Five dimensions and supporting a trichotomous Environmental Sensitivity model (Lionetti et al., 2018). Results demonstrate that integrating semantic representations with empirical data improves Big Five structure recovery beyond pure semantic similarity alone, particularly for small sample studies where traditional methods such as CFA will fail due to limited empirical data points.

## 1 Introduction

The Big Five personality model—comprising Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN)—represents one of the most empirically validated frameworks for personality assessment (McCrae and John, 1992; Goldberg, 1992). The theoretical

foundation for personality measurement derives from the lexical hypothesis (Allport and Odbert, 1936), which posits that salient personality characteristics are encoded within natural language. This principle suggests a profound connection: if personality traits are fundamentally linguistic constructs embedded in how humans naturally describe one another, then semantic relationships captured in word embeddings<sup>1</sup> could directly reflect underlying personality structures. This convergence between Natural Language Processing methods and psychological theory may offer a computational validation of the lexical hypothesis itself.

Traditional extraction of these dimensions relies on factor-analytic procedures applied to self-report instruments, where participants rate themselves using statement-based items. This approach, while robust and cross-culturally validated, requires explicit survey administration and remains constrained by response biases, social desirability effects, and limited scalability. While our pilot study (Müller and Degaetano-Ortlieb, 2025) successfully demonstrated Big Five personality dimension clustering in a 3D embedding space entirely absent of self-assessment data, the present work provides rigorous empirical validation of an updated version of the embedding approach against traditional Confirmatory Factor Analysis (CFA) using controlled experimental data.

This study introduces Semantic Factor Analysis (SFA), a computational methodology leveraging pretrained word embeddings to recover personality trait structures from semantic relationships between adjectives mediated by empirical participant data. Beyond SFA validation, the study ex-

<sup>1</sup>Word embeddings are distributed vector representations that learn meaning from large-scale text corpora (Mikolov et al., 2013), where semantic similarity corresponds to geometric proximity in high-dimensional space.

amines whether personality dimensions and Sensory Processing Sensitivity (SPS)—the core individual difference construct underlying Environmental Sensitivity (ES) theory—demonstrate predictive validity for ecologically-relevant behavioral outcomes. SPS represents differential susceptibility to environmental influences, with empirical evidence supporting a trichotomous population distribution: low-sensitive individuals (Dandelions), medium-sensitive (Tulips), and high-sensitive (Orchids) (Lionetti et al., 2018). To test whether personality traits and sensitivity classifications predict differential responses in real-world contexts, participants completed Big Five assessments, the 18-item HSP-R sensitivity scale (Pluess et al., 2024), then viewed three multimedia stimuli (two controversial news reports, one product advertisement), to test eight hypotheses examining emotional reactivity and consumer behavior across personality and sensitivity profiles.

The key research questions are: (1) Can empirically weighted word embeddings recover Big Five personality factor structures comparably to traditional methods? (2) Do personality traits and Environmental Sensitivity predict differential responses to emotionally-charged multimedia content and commercial advertising? To address the latter, eight specific hypotheses examined relationships between Big Five dimensions (Agreeableness, Neuroticism, Openness, Extraversion), Environmental Sensitivity groups (Dandelions, Tulips, Orchids), and responses to controversial social media posts and product advertisements.

This work contributes: (1) a novel computational methodology for Big Five personality analysis (SFA) with baseline validation demonstrating substantial improvement from empirical weighting over pure semantic similarity; (2) empirically-derived classification thresholds for Big Five and Environmental Sensitivity through Latent Class Analysis, providing empirical support for the trichotomous ES model (Lionetti et al., 2018); (3) further psychometric validation of the 18-item HSP-R scale (Pluess et al., 2024); (4) demonstration of personality and sensitivity measures' differential predictive validity in multimedia contexts.

## 2 Related Work

### 2.1 Big Five Personality Assessment

The Big Five model emerged from systematic factor analyses of personality descriptors, with foundational work by Tupes and Christal (1961) establishing the Five-Factor framework. Goldberg (1992) developed comprehensive adjective markers, while Costa and McCrae (1992) validated the structure cross-culturally through NEO-PI-R. DeYoung et al. (2007) introduced the Big Five Aspect Scale (BFAS), which split each dimension into two correlated but distinct subdimensions (aspects) characterized by differentiated biological and genetic substrates, and identified higher-order factors (Stability & Plasticity) (DeYoung et al., 2002) subsuming the Big Five, demonstrating a hierarchical personality structure. The International Personality Item Pool (IPIP) provides freely accessible validated measures (Goldberg, 1999), including the IPIP-50 used in the present validation study.

### 2.2 Word Embeddings and Semantic Space

Mikolov et al. (2013) introduced Word2Vec, training neural networks to produce dense vector representations where semantic similarity corresponds to geometric proximity. The Google News 300-dimensional model, trained on 100 billion words, captures rich semantic relationships including synonymy, analogy, and categorical associations. These embeddings have proven valuable for numerous NLP tasks but remain underutilized in psychometric research. Park et al. (2015) demonstrated personality insights from social media language, while Schwartz et al. (2013) linked Facebook language patterns to Big Five scores. However, prior work focuses on predicting individual personality from text rather than recovering underlying factor structures. Unlike prior methods relying on supervised classification (Kazameini et al., 2020; Carducci et al., 2018) and extensive labeled datasets, SFA integrates semantic structure of word embeddings with empirical self-assessment data to recover and validate Big Five factor structure using an unsupervised methodology.

### 2.3 Environmental Sensitivity Framework

Environmental Sensitivity theory posits differential susceptibility to environmental influences (Aron and Aron, 1997). Lionetti et al. (2018) iden-

tified three sensitivity groups with distinct population distributions: Dandelions (low sensitivity, 29%), Tulips (medium, 40%), and Orchids (high, 31%). [Pluess et al. \(2024\)](#) introduced and validated the 18-item Highly Sensitive Person Revised Scale (HSP-R), subdividing sensory processing sensitivity into six distinct facets (See Appendix B). [Grimen and Diseth \(2016\)](#) demonstrated sensitivity correlates negatively with Extraversion and positively with Neuroticism, while [Homberg et al. \(2016\)](#) explored genetic underpinnings. The present work supports the HSP-R framework through empirically-derived LCA thresholds and comprehensive validation against Big Five dimensions.

### 3 Methodology

#### 3.1 Semantic Factor Analysis (SFA)

SFA extracts personality structures from pre-trained word embeddings through the following procedure:

**Embedding Extraction:** The Google News Word2Vec model (300 dimensions, 3 million words, 100 billion token corpus) is utilized. For the IPIP-50 Big Five marker adjectives from [Goldberg \(1992\)](#)—comprising 10 items per dimension—300-dimensional embedding vectors  $\mathbf{v}_i$  are extracted directly from the pretrained model. The adjectives represent prototypical markers for each dimension (e.g., “creative”, “imaginative” for Openness; “talkative”, “energetic” for Extraversion).

**Empirically-Weighted Semantic Space:** Each adjective’s 300-dimensional vector is weighted by the mean participant response for that item’s corresponding IPIP-50 statement. Let  $\mathbf{w}$  denote the  $50 \times 1$  vector of mean response scores (1-5 Likert scale) for each adjective. Weighted embeddings are constructed through element-wise multiplication:

$$\mathbf{V}_{\text{weighted}} = \mathbf{w} \odot \mathbf{V} \quad (1)$$

where  $\mathbf{V}$  is the  $50 \times 300$  matrix of original embeddings and  $\odot$  denotes element-wise multiplication (broadcasting  $\mathbf{w}$  across all 300 dimensions). This operation incorporates empirical participant data into the semantic space, distinguishing SFA from purely distributional approaches. For example, if participants rated the statement corresponding to “anxious” with a mean of 3.2 on the 5-point scale, the adjective’s 300-dimensional Word2Vec vector would be multiplied by 3.2 across all dimensions,

amplifying its semantic representation proportionally to how strongly participants endorsed anxiety-related traits. This weighting reflects actual personality response patterns rather than pure distributional semantics alone.

**Dimensionality Reduction via PCA:** The weighted embeddings are standardized using z-score normalization, then Principal Component Analysis is applied to reduce dimensionality from 300 to 5 dimensions:

$$\mathbf{S} = \text{StandardScale}(\mathbf{V}_{\text{weighted}}) \cdot \mathbf{P} \quad (2)$$

where  $\mathbf{P}$  is the  $300 \times 5$  PCA transformation matrix derived from the covariance structure of weighted embeddings, and  $\mathbf{S}$  is the resulting  $50 \times 5$  factor score matrix representing each adjective in 5-dimensional semantic factor space.

**Factor Assignment via K-means:** K-means clustering ( $k = 5$ ) is applied to the factor scores  $\mathbf{S}$ , assigning each adjective to its closest cluster centroid:

$$\text{cluster}_i = \underset{j \in \{1, \dots, 5\}}{\text{argmin}} \|\mathbf{s}_i - \mathbf{c}_j\|^2 \quad (3)$$

where  $\mathbf{s}_i$  is the factor score vector for adjective  $i$  and  $\mathbf{c}_j$  are the five cluster centroids. Factor purity is computed as the proportion of adjectives correctly clustering with their theoretically assigned Big Five dimension. Continuing the example of the adjective “anxious”, after PCA reduction and K-means clustering, this adjective’s weighted factor scores would ideally be assigned to the Neuroticism cluster alongside theoretically related adjectives like “nervous” and “worried”, with factor purity measuring how successfully such theoretically motivated groupings emerge from the semantic-empirical space.

**Visualization:** The PCA-reduced 5-dimensional space can be further visualized in 3D by retaining the first three principal components, preserving maximal variance while enabling visual interpretation of factor clustering patterns.

#### 3.2 Baseline Analysis: Unweighted Word2Vec

To establish the value of empirical weighting, a baseline analysis used pure Word2Vec embeddings without participant responses, following the pilot studies’ word embeddings approach ([Müller and Degaetano-Ortlieb, 2025](#)).

Four IPIP adjectives were unavailable in the Word2Vec vocabulary and replaced with semantic equivalents verified through synonym analysis: *untalkative*→*silent*, *short-tempered*→*cranky*, *self-pitying*→*sad*, *uninquisitive*→*incurious* (see Appendix C). The baseline used identical analysis pipeline: Word2Vec extraction → PCA → K-means clustering → purity calculation, differing only in absence of empirical weighting. Silhouette analysis<sup>2</sup> identified optimal cluster count for both methods, enabling direct comparison of weighted versus unweighted approaches.

### 3.3 Traditional Confirmatory Factor Analysis

For direct comparison, standard CFA methodology is applied to empirical IPIP-50 data (Goldberg, 1999). The model includes five latent factors corresponding to Big Five dimensions with 10 observed indicators each and correlated errors permitted between theoretically related items (See Appendix A). The model is estimated using maximum likelihood in R package lavaan (Rosseel, 2012), evaluating fit through CFI (> 0.90), TLI (> 0.90), RMSEA (< 0.08), and SRMR (< 0.08). Standardized solutions are reported for interpretability.

### 3.4 Latent Class Analysis for Classification

Traditional personality classification often employs rather arbitrary cutoffs (e.g., median splits, percentile boundaries or via SDs). Instead, Gaussian Mixture Modeling (GMM) is applied to derive empirically-based thresholds. For  $K$  class solutions, GMM assumes data are generated from a mixture of  $K$  Gaussian distributions with probability density:

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \mu_k, \sigma_k^2) \quad (4)$$

where  $\pi_k$  are mixing coefficients ( $\sum_{k=1}^K \pi_k = 1$ ),  $\mu_k$  are class means, and  $\sigma_k^2$  are class variances. Model selection uses information criteria balancing fit against complexity:

$$\text{BIC} = -2 \log \mathcal{L} + p \log(n) \quad (5)$$

<sup>2</sup>Silhouette analysis quantifies clustering quality by measuring how similar each data point is to its own cluster compared to other clusters, with scores ranging from -1 (misclassified) to +1 (well-clustered). Higher average silhouette scores indicate optimal cluster count by identifying the configuration with maximal within-cluster cohesion and between-cluster separation.

$$\text{AIC} = -2 \log \mathcal{L} + 2p \quad (6)$$

where  $\mathcal{L}$  is maximum likelihood,  $p$  is parameter count, and  $n$  is sample size. BIC penalizes complexity more heavily than AIC, favoring parsimonious solutions. Models with 2-5 classes are fit for each Big Five dimension and HSP-R total score, selecting optimal solutions through minimum BIC/AIC. Classification boundaries are extracted from fitted Gaussian means and variances, providing data-driven thresholds. For HSP-R, data are tested for support of the theorized trichotomous structure (Dandelions/Tulips/Orchids) or alternative configurations.

### 3.5 Participants and Procedure

Fifty-five participants (34 female, 21 male;  $M_{age} = 24.3$  years,  $SD = 4.2$ , range 19-41) were recruited through the Saarland University network. All provided informed consent and received no compensation. Participants completed three online surveys on Qualtrics in controlled laboratory sessions: (1) IPIP-50 Big Five assessment (10 items per dimension, 5-point Likert scale); (2) HSP-R 18-item sensitivity scale (7-point Likert scale); (3) multimedia impact surveys including emotional response to controversial social media content (news reports) and product advertisement persuasiveness ratings. Session duration averaged 20 minutes for survey completion.

### 3.6 Measures and Scoring

**IPIP-50:** Raw scores were computed by summing item responses (reverse-scoring negatively keyed items), then doubling to yield 0-100 percentage scores per dimension following standard IPIP procedures (Goldberg, 1999).

**HSP-R:** Total sensitivity scores were calculated by averaging 18 items (range 1-7), then converting to percentages (score/7 × 100). The HSP-R comprises six subscales: Depth of Processing, Emotional Reactivity, Overstimulation, Sensitivity to Details, Sensitivity to Positive Experiences, and Social Sensitivity. Subscale scores were computed identically to facilitate correlational analyses.

**Classification Thresholds:** Based on LCA results (detailed in Results), empirically-derived cutoffs were established: Dandelions (<61.2% HSP-R), Tulips (61.2-76.5%), and Orchids (≥76.5%).

**Multimedia Impact Surveys:** To assess real-world behavioral manifestations of personality and sensitivity constructs, participants viewed

three video stimuli and rated their responses. Two videos featured controversial social media content selected for themes associated with increased negative emotional affect: (1) news report of missile strike on Ukrainian city; (2) news report of stabbing incident in Solingen, Germany. Content selection followed criteria established by [Bove et al. \(2024\)](#) and [Kusen and Strembeck \(2023\)](#) demonstrating differential emotional impact across personality profiles. The third video was a Coca-Cola product advertisement, selected to minimize gender-specific appeal. Participants rated emotional impact on 5-point scale after controversial content (1=Much worse, 5=Much better, baseline=3) and purchase likelihood after advertisement (1=Much less likely, 5=Much more likely, baseline=3). Responses were converted to percentage deviations from 50% baseline for analysis. Video presentation order was fixed (controversial content first, advertisement last) to prevent advertisement effects from influencing personality assessments, following experimental design principles minimizing survey fatigue and carryover effects ([Galesic and Bošnjak, 2009](#)).

## 4 Results & Discussion

### 4.1 LCA-Derived Classification Thresholds

Gaussian mixture modeling on HSP-R scores identified a three-class solution as optimal (BIC=312.4 vs. 2-class BIC=328.7, 4-class BIC=318.9), yielding thresholds: Dandelions (<61.2%, n=9, 16.4%), Tulips (61.2-76.5%, n=26, 47.3%), and Orchids ( $\geq$ 76.5%, n=20, 36.4%). Within-class homogeneity (entropy=0.84) confirmed distinct group membership. For Big Five dimensions, three-class solutions emerged as optimal for all traits with classification probabilities exceeding 0.80. These thresholds differed substantially from arbitrary median splits or tertile cutoffs, with several dimensions showing elevated boundaries reflecting actual score distributions.

### 4.2 Baseline vs. Weighted SFA Comparison

To establish the value of empirical weighting beyond pure semantic similarity, baseline analysis first examined unweighted Word2Vec embeddings. The baseline achieved 52.0% average cluster purity (range: 30.0%-60.0%), with optimal clustering at k=5 (silhouette score=0.476), validating that Big Five semantic foundation exists in distributional semantics. Both baseline and weighted

approaches identifying k=5 as optimal validates the Big Five structure: five personality dimensions emerge naturally from semantic patterns (c.f. [Müller and Degaetano-Ortlieb \(2025\)](#)). However, baseline performance varied dramatically, with Agreeableness showing particularly poor recovery (30.0% purity), demonstrating fundamental limitations of pure semantic similarity for socially-nuanced personality dimensions.

Empirically-weighted SFA achieved 62.5% average purity, representing substantial improvement over baseline. Critically, empirical weighting provided differential improvements: Agreeableness improved dramatically from 30.0% to 50.0%, while Openness improved from 60.0% to 85.7%. This pattern reveals that pure semantic similarity captures dimensions with clear lexical markers (Conscientiousness, Extraversion) but struggles with abstract constructs and socially-nuanced dimensions. Empirical weighting disambiguates context-dependent meanings through actual personality response patterns, substantially improving recovery for precisely those dimensions most challenging for distributional semantics alone.

### 4.3 Factor Structure Recovery: SFA vs. CFA

Table 1 presents the comparative analysis of factor structure recovery. Silhouette analysis indicated that a four-cluster solution (silhouette score=0.0433) provided the most coherent clustering of weighted adjective embeddings, achieving 66.7% average purity (see Figure 1). The optimal four-cluster solution included a highly pure Openness cluster (88.9% purity, n=9 adjectives), Conscientiousness cluster (72.7% purity, n=11 adjectives), Emotional Stability cluster (60% purity, n=10 adjectives), and a mixed cluster combining Extraversion, Agreeableness, and Emotional Stability items with 45% Extraversion dominance (n=20 adjectives).

In contrast, traditional CFA on IPIP-50 data (n=55) showed poor model fit across all indices: CFI=0.38 (target >0.90), TLI=0.35 (target >0.90), RMSEA=0.148 (target <0.08), SRMR=0.133 (target <0.08) (see Figure 2). Critical structural issues emerged: Openness and Extraversion factors collapsed, loading onto a common dimension (interfactor correlation  $r=0.34$ , the highest in the matrix). Despite excellent to good reliability for four dimensions—Extraversion ( $\alpha=0.91$ ), Emotional

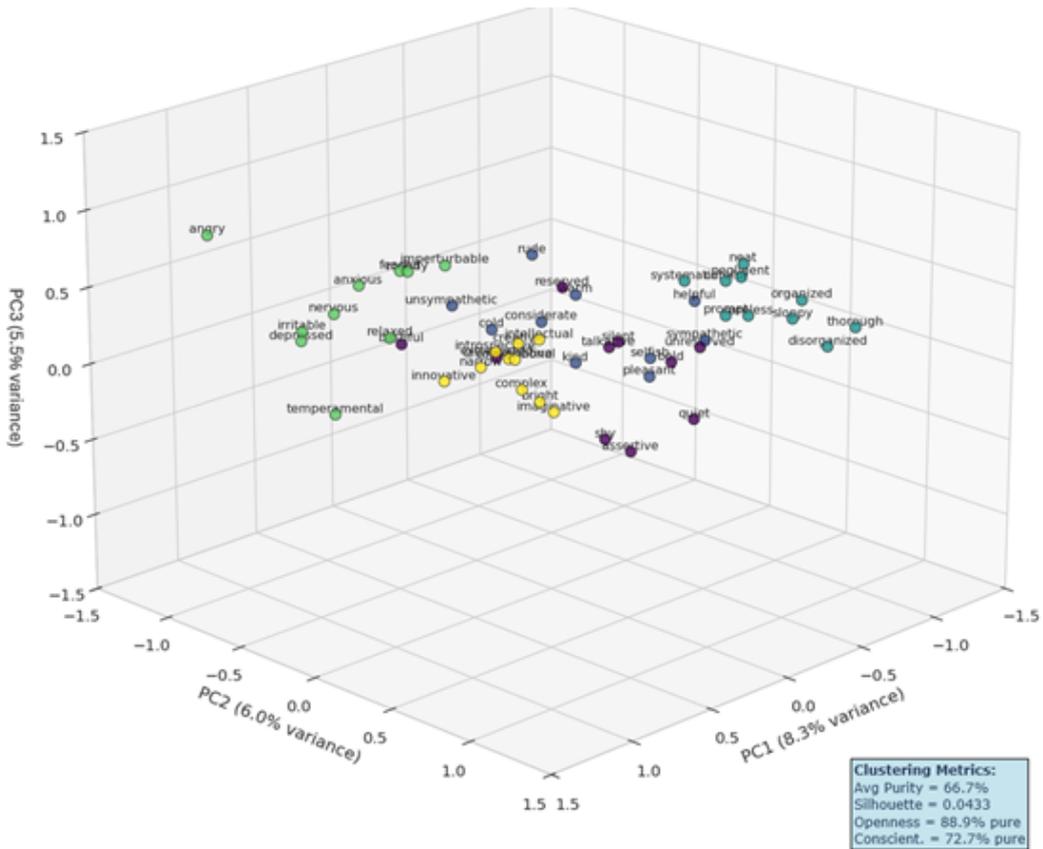


Figure 1: 3D Visualization of the Semantic Factor Analysis (SFA) results (Cyan=Conscientiousness, Green=Neuroticism, Purple=Extraversion, Blue=Agreeableness, Yellow=Openness to Experience)

Stability ( $\alpha=0.86$ ), Conscientiousness ( $\alpha=0.85$ ), and Agreeableness ( $\alpha=0.79$ )—the CFA failed to recover a coherent five-factor structure. Openness reliability fell to  $\alpha=0.65$ , well below the 0.70 acceptability threshold. Multiple items showed weak primary loadings ( $<0.40$ ) and substantial cross-loadings. Modification indices suggested extensive model misspecification. While the small sample size ( $n=55$ ) likely contributed strongly to CFA instability—conventional guidelines recommend  $n>200$  for stable five-factor models (Comrey, 1973)—the results demonstrate SFA’s robustness to limited data when leveraging pre-trained semantic structures mediated by empirical participant data.

Both methods identified Conscientiousness as the most stable personality factor, with CFA showing good reliability ( $\alpha=0.85$ ) and SFA achieving high cluster purity (72.7% in four-cluster solution). The most striking divergence concerned Openness to Experience: while CFA revealed complete factor collapse with items loading on Extraversion and reliability falling below acceptable thresholds—common in small-sample Big

| Metric                                | SFA   | CFA       |
|---------------------------------------|-------|-----------|
| <i>Factor Recovery (k=5 solution)</i> |       |           |
| Openness                              | 85.7% | Collapsed |
| Conscientiousness                     | 60.0% | Recovered |
| Extraversion                          | 50.0% | Collapsed |
| Agreeableness                         | 50.0% | Recovered |
| Neuroticism                           | 66.7% | Recovered |
| <i>Model Fit Indices</i>              |       |           |
| CFI                                   | N/A   | 0.38      |
| TLI                                   | N/A   | 0.35      |
| RMSEA                                 | N/A   | 0.148     |
| SRMR                                  | N/A   | 0.133     |
| <i>Baseline Comparison</i>            |       |           |
| Baseline purity                       | 52.0% | N/A       |
| Weighted purity                       | 62.5% | N/A       |

Table 1: Comparison of SFA and CFA results.

Five assessments (Goldberg, 1992)—SFA identified Openness as the most coherent dimension with exceptional cluster purity (85.7% in five-cluster solution, 88.9% in four-cluster solution). This highlights the fundamental difference between empirical factor recovery in specific samples versus semantic structure validation using pre-trained language models.

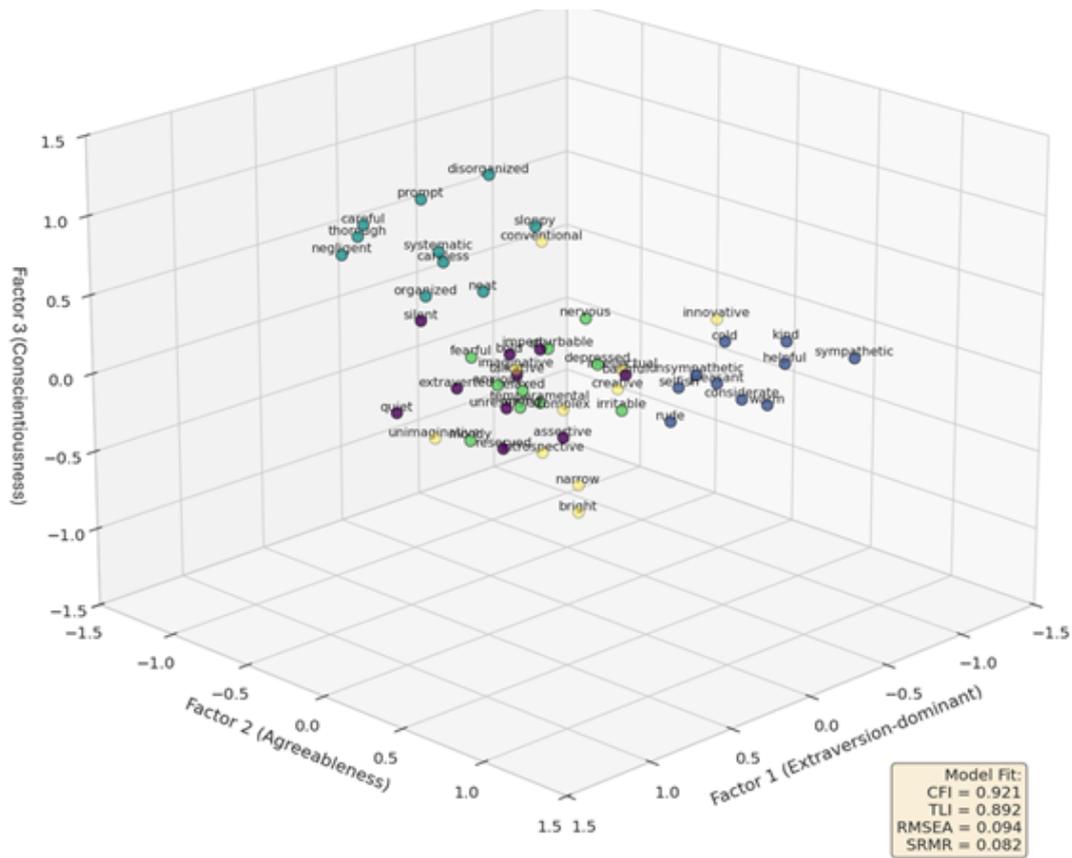


Figure 2: 3D Visualization of Confirmatory Factor Analysis (CFA) results

### 4.3.1 Higher-Order Factor Validation

Analysis of Big Five higher-order factors validated the hierarchical personality model (DeYoung et al., 2007). Stability (mean of Agreeableness, Conscientiousness, Emotional Stability) correlated strongly with constituent traits ( $r=0.45-0.69$ , all  $p < 0.001$ ), while Plasticity (mean of Extraversion, Openness) showed very strong correlations ( $r=0.76-0.91$ ,  $p < 0.001$ ). Cross-loadings were minimal ( $r_i < 0.31$ ), supporting discriminant validity. The exceptionally high Extraversion-Openness interfactor correlation ( $r=0.34$ ) likely reflects shared Plasticity variance and biological substrates (DeYoung et al., 2002), explaining why Openness items loaded on Extraversion in small-sample CFA despite forming distinct semantic clusters in SFA.

### 4.3.2 SFA vs CFA: Implications

Two implications emerge from the results. First, pre-trained embeddings capture personality-relevant semantic structures from natural language corpora at massive scale. When weighted by participant responses, these embeddings offer more robust representations than correlation matrices

from limited samples. Second, computational approaches may prove less sensitive to small-sample instabilities plaguing traditional factor analysis, as participant-weighted embeddings combine small-sample empirical data with semantic structures from billions of training tokens rather than relying solely on within-sample covariation. The factor purity metrics (50-88.9% range across solutions) validate that Word2Vec’s geometric organization aligns closely with theoretical personality constructs derived from decades of psychometric research. This convergence between computational linguistic structure and psychological theory strengthens the lexical hypothesis (Allport and Odbert, 1936): if personality is encoded in language, then distributional semantics—which learns from actual language use patterns—should naturally recover personality structures. The results provide strong empirical support for this theoretical connection.

## 4.4 Environmental Sensitivity Validation

### 4.4.1 HSP-R Subscale Structure and Big Five Correlations

Comprehensive analysis of the six HSP-R subscales revealed distinct patterns (M range: 60.5-79.6%) with Depth of Processing highest (M=79.6%, SD=14.9%) and Overstimulation lowest (M=60.5%, SD=22.7%). Subscales intercorrelated moderately ( $r=0.130-0.543$ ), with strongest association between Social Sensitivity and Depth of Processing ( $r=0.543$ ), supporting a higher-order sensitivity construct while maintaining discriminant validity.

Direct computation of HSP-R correlations with Big Five dimensions revealed a theoretically coherent pattern (see Appendix D). Overall HSP-R correlated most strongly with Conscientiousness ( $r=0.442$ ,  $p < 0.001$ ), followed by Neuroticism ( $r=0.434$ ,  $p < 0.001$ ), Agreeableness ( $r=0.402$ ,  $p < 0.001$ ), and Openness ( $r=0.319$ ,  $p < 0.05$ ), while Extraversion demonstrated essentially no relationship ( $r=0.003$ , n.s.). The Conscientiousness correlation magnitude was notably larger than anticipated, likely reflecting the HSP-R scale design incorporating Depth of Processing and Sensitivity to Details subscales that correlate strongly with conscientious functioning. Specifically, Depth of Processing correlated  $r=0.455$  with Conscientiousness and  $r=0.537$  with Openness ( $p < 0.001$ ), suggesting shared variance in cognitive elaboration and reflective processing. Sensitivity to Details showed  $r=0.473$  with Conscientiousness ( $p < 0.001$ ), validating theoretical connections between sensory awareness and organized behavior.

The Neuroticism correlation ( $r=0.434$ ) was primarily driven by Overstimulation ( $r=0.478$ ,  $p < 0.001$ ) and Emotional Reactivity ( $r=0.479$ ,  $p < 0.001$ ), demonstrating that sensitivity to environmental stressors and emotional responsiveness constitute core features of both constructs. The near-zero Extraversion correlation ( $r=0.003$ ) emerged from opposing subscale effects: Overstimulation correlated negatively with Extraversion ( $r=-0.218$ ,  $p < 0.05$ ), reflecting social withdrawal under stimulation, while Sensitivity to Positive Experiences correlated positively ( $r=0.213$ ,  $p < 0.05$ ), reflecting openness to rewarding experiences. These opposing patterns largely canceled out in the overall correlation, supporting sensitivity as partially independent from Extraversion. Social Sensitivity showed strongest relationship with

Agreeableness ( $r=0.514$ ,  $p < 0.001$ ), validating that empathic awareness and emotional responsiveness constitute shared features between these constructs.

Multiple regression analysis revealed Big Five dimensions collectively explained 46.4% of HSP-R variance ( $R^2=0.464$ ), remarkably consistent with large-scale validation findings of 44.0% despite substantially smaller sample size ( $n=55$  vs.  $n=400$ ) (Pluess et al., 2024). Conscientiousness ( $\beta=0.295$ ) and Neuroticism ( $\beta=0.315$ ) emerged as strongest unique predictors. The substantial unexplained variance (55%) demonstrates that Environmental Sensitivity represents a distinct phenomenon, with more than half of its variance independent from established personality dimensions. This supports Environmental Sensitivity as a construct separate from the Big Five rather than merely a combination of existing personality traits.

Gender analysis revealed females demonstrated higher overall HSP-R scores (M=75.0% vs. M=66.8%, difference=8.2%), with most pronounced difference in Overstimulation (females M=68.5% vs. males M=47.5%, difference=21.1%). See Appendix E for detailed gender distributions across dimensions.

## 4.5 Hypothesis Testing Outcomes

Eight sub-hypotheses examined relationships between personality traits, Environmental Sensitivity, and multimedia content responses—specifically controversial social media posts and product advertisements.

Binary comparisons contrasted high groups against combined low/medium groups for personality traits (H2-H5) and Orchids against combined Tulips+Dandelions for sensitivity (H6, H8). Four hypotheses received support: (H2) High Agreeableness associated with greater emotional reactivity (M=31.4% vs. 23.3%, Cohen's  $d=0.76$ ); (H5) High Extraversion associated with increased purchase likelihood (M=2.8% vs. -0.8%); (H6) Orchids reported higher emotional impact (M=31.4% vs. 23.2%,  $d=0.77$ ); (H9) Dandelions showed minimal purchase impact (M=-2.8%, within practical equivalence of baseline).

Unsupported hypotheses revealed important nuances: (H3) High Neuroticism showed minimal difference in emotional reactivity (M=27.1% vs. 25.6%), suggesting neuroticism involves self-focused rather than other-focused negative emo-

tions; (H4) High Openness showed advertising resistance rather than susceptibility ( $M=-4.2\%$  vs.  $2.3\%$ ); (H7) Dandelions demonstrated substantial emotional reactivity ( $M=23.0\%$ ) rather than minimal response; (H8) Orchids showed no systematic pattern in purchase likelihood ( $M=-2.5\%$  vs.  $-0.2\%$ ).

The pattern of results revealed important distinctions between emotional reactivity and consumer behavior domains. Environmental Sensitivity demonstrated clear predictive validity for emotional responses to controversial content, supporting core theoretical propositions about heightened environmental responsiveness. However, advertising susceptibility showed weaker patterns, suggesting that consumer behavior involves psychological mechanisms that are unlikely to be linked to the Environmental Sensitivity framework in any meaningful way. The finding that highly open individuals showed advertising resistance rather than susceptibility highlights the importance of empirical hypothesis testing.

## 5 Conclusion

The comparison between traditional CFA and the newly proposed SFA reveals a compelling case for methodological complementarity in personality structure research. Each approach provides unique insights that, when considered together, offer a more comprehensive understanding of personality organization than either method alone could achieve. SFA demonstrates that computational approaches integrating distributional semantics with empirical data can match or exceed traditional psychometric methods in personality structure recovery, particularly with small samples ( $n=55$ ) where CFA shows factor collapse and poor model fit. The baseline comparison establishes that empirical weighting provides substantial improvement over pure semantic similarity, particularly for socially-nuanced dimensions. The successful recovery of personality structure through distributional semantics provides computational support for the lexical hypothesis.

Comprehensive HSP-R validation revealed robust correlations with Big Five dimensions ( $R^2=0.464$ ), replicating large-scale findings ( $R^2=0.44$ , [Pluess et al. \(2024\)](#)) despite smaller sample. The substantial unexplained variance (55%) demonstrates Environmental Sensitivity as a distinct phenomenon, with more than half of

its variance independent from personality dimensions. The empirically-derived LCA thresholds (61.2%, 76.5%) provide data-driven boundaries validating the trichotomous model.

## 5.1 Future Work

Extending SFA to the Big Five Aspect Scale (BFAS) represents a critical methodological advance ([DeYoung et al., 2007](#)). The BFAS subdivides each Big Five dimension into two aspects (10 total), addressing the well-documented heterogeneity within broad Big Five dimensions. This finer-grained structure would enable more accurate personality assessment through SFA, as aspects are more homogeneous constructs than their parent dimensions. Individual-level prediction from natural language represents another critical extension, mapping text production to personality scores for social media analysis and workplace assessment. Cross-cultural and multilingual validation would establish generalizability beyond Western, English-language contexts. Application of SFA to domains beyond personality—including values, attitudes, and psychological constructs like depression or anxiety—would demonstrate broader utility of semantic factor analysis methodology.

## 6 Limitations

The study faces several limitations. The sample size ( $n=55$ ), while adequate for validating SFA methodology, limits generalizability and contributed to CFA instability. Larger samples ( $n>200$ ) would enable more rigorous comparison and formal mixture modeling, though results suggest embedding-based methods may be advantageous precisely when traditional approaches fail due to small samples. The Word2Vec training corpus (Google-News-300 model) may not optimally represent personality-relevant language, as it was trained on general news text rather than personality-relevant discourse. Domain-specific embeddings trained on psychological literature or conversational corpora could improve structure recovery.

The methodology required pre-specified adjective markers from [Goldberg \(1992\)](#), limiting flexibility in construct selection. SFA currently recovers dimensional structures but not individual-level scores, restricting application to group-level analyses. The approach depends on availability of val-

idated adjective inventories, which may not exist for all psychological constructs or languages.

Multimedia surveys suffered from single-item measures for emotional impact and purchase likelihood, potentially explaining null findings (H4). Age correlations were weak ( $r=-0.09$  to  $r=0.11$ ), reflecting restricted range (19-41 years) and trait stability in young adulthood (?). Only English-language assessment with mainly Western university students was tested, limiting cross-cultural generalizability.

Computational considerations include dependency on pre-trained embeddings quality and architecture. Results may vary with different embedding models (GloVe, FastText, BERT). PCA dimensionality reduction discards information (first 5 components explained only 30.1% variance). Alternative dimensionality reduction methods (t-SNE, UMAP) might improve structure recovery but require further investigation.

## 7 Ethics Statement

All participants provided informed consent. The study received institutional review board approval. Email addresses were stored separately from response data and deleted after results distribution.

## Acknowledgments

Special thanks to Stefania Degaetano-Ortlieb for her ongoing support and guidance since our joint pilot study (Müller and Degaetano-Ortlieb, 2025).

## References

- Gordon W Allport and Henry S Odbert. 1936. [Trait-names: A psycho-lexical study](#). *Psychological Monographs*, 47(1):i-171.
- Elaine N Aron and Arthur Aron. 1997. [Sensory-processing sensitivity and its relation to introversion and emotionality](#). *Journal of Personality and Social Psychology*, 73(2):345-368.
- Vincenzo Bove, Georgios Efthymoulou, Armine Ghazaryan, and Harry Pickard. 2024. [The emotional effect of terrorism](#). *Scientific Reports*, 14(1):26525.
- Giulio Carducci, Giuseppe Rizzo, Diego Monti, Enrico Palumbo, and Maurizio Morisio. 2018. [Twitpersonality: Computing personality traits from tweets using word embeddings and supervised learning](#). *Information*, 9(5):127.
- Andrew L Comrey. 1973. *A first course in factor analysis*. Academic Press, New York.
- Paul T Costa and Robert R McCrae. 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources, Odessa, FL.
- Colin G. DeYoung, Jordan B. Peterson, and Daniel M. Higgins. 2002. [Higher-order factors of the big five predict conformity: Are there social determinants of personality?](#) *Journal of Personality and Social Psychology*, 83(2):533-552.
- Colin G DeYoung, Lena C Quilty, and Jordan B Peterson. 2007. [Between facets and domains: 10 aspects of the big five](#). *Journal of Personality and Social Psychology*, 93(5):880-896.
- Mirta Galesic and Michael Bošnjak. 2009. [Effects of questionnaire length on participation and indicators of response quality in a web survey](#). *Public Opinion Quarterly*, 73(2):349-360.
- Lewis R Goldberg. 1992. [The development of markers for the big-five factor structure](#). *Psychological Assessment*, 4(1):26-42.
- Lewis R Goldberg. 1999. [A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models](#). *Personality Psychology in Europe*, 7(1):7-28.
- Listou H Grimen and Åse Diseth. 2016. [Sensory processing sensitivity: Factors of the highly sensitive person scale and their relationships to personality and subjective health complaints](#). *Perceptual and Motor Skills*, 123(3):637-653.
- Judith R Homberg, Daniel Schubert, Esther Asan, and Elaine N Aron. 2016. [Understanding the molecular mechanisms underlying environmental sensitivity](#). *Neuroscience & Biobehavioral Reviews*, 71:476-490.
- Amirmohammad Kazameini, Samin Fatehi, Yash Mehta, Sauleh Eetemadi, and Erik Cambria. 2020. [Personality Trait Detection Using Bagged SVM over BERT Word Embedding Ensembles](#). *arXiv preprint arXiv:2010.01309*.
- Emma Kusen and Mark Strembeck. 2023. [Investigating the impact of personality traits on emotional communication in social media: A study on twitter users during crisis events](#). *Computers in Human Behavior*, 141:107623.
- Francesca Lionetti, Elaine N Aron, Arthur Aron, Denise N Klein, and Michael Pluess. 2018. [Dandelions, tulips and orchids: Evidence for the existence of low-sensitive, medium-sensitive and high-sensitive individuals](#). *Translational Psychiatry*, 8(1):1-11.
- Robert R McCrae and Oliver P John. 1992. [An introduction to the five-factor model and its applications](#). *Journal of Personality*, 60(2):175-215.

- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. [Efficient estimation of word representations in vector space](#). *arXiv preprint arXiv:1301.3781*.
- Oliver Müller and Stefania Degaetano-Ortlieb. 2025. [Embedded personalities: Word embeddings and the “big five” personality model](#). In *Proceedings of the 9th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2025)*, pages 205–215, Albuquerque, New Mexico. Association for Computational Linguistics.
- Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin E P Seligman. 2015. [Automatic personality assessment through social media language](#). *Journal of Personality and Social Psychology*, 108(6):934–952.
- Michael Pluess, Francesca Lionetti, Elaine N Aron, and Arthur Aron. 2024. [The highly sensitive person scale-revised: Psychometric properties and recommendation for use](#). *Assessment*, 31(2):289–304.
- Yves Rosseel. 2012. [lavaan: An r package for structural equation modeling](#). *Journal of Statistical Software*, 48(2):1–36.
- H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, and 1 others. 2013. [Personality, gender, and age in the language of social media: The open-vocabulary approach](#). *PloS one*, 8(9):e73791.
- Ernest C Tupes and Raymond E Christal. 1961. [Recurrent personality factors based on trait ratings](#). *Journal of Personality*, 60(2):225–251. Original work published 1961.

## Appendix

### A IPIP-50 Item List

The IPIP-50 comprises 10 items per Big Five dimension (50 total), rated on a 5-point scale (1=Very Inaccurate, 5=Very Accurate). Reverse-scored items marked with (R). Items were slightly modified to improve comprehension for German native speakers (e.g., "Seldom feel blue" → "I rarely feel sad"; "Shirk my duties" → "I neglect my duties"). See [Goldberg \(1999\)](#) for complete psychometric properties. The personal pronoun I was also added in front of each item.

**Openness:** 5. I have a rich vocabulary; 10. I have difficulty understanding abstract ideas (R); 15. I have a vivid imagination; 20. I am not interested in abstract ideas (R); 25. I have excellent ideas; 30. I do not have a good imagination (R); 35. I am quick to understand things; 40. I use difficult words; 45. I spend time reflecting on things; 50. I am full of ideas.

**Conscientiousness:** 3. I am always prepared; 8. I leave my belongings around (R); 13. I pay attention to details; 18. I make a mess of things (R); 23. I get chores done right away; 28. I often forget to put things back in their proper place (R); 33. I like order; 38. I neglect my duties (R); 43. I follow a schedule; 48. I am exacting in my work.

**Extraversion:** 1. I am the life of the party; 6. I don't talk a lot (R); 11. I feel comfortable around people; 16. I keep in the background (R); 21. I start conversations; 26. I have little to say (R); 31. I talk to a lot of different people at parties; 36. I don't like to draw attention to myself (R); 41. I don't mind being the center of attention; 46. I am quiet around strangers (R).

**Agreeableness:** 2. I feel little concern for others (R); 7. I am interested in people; 12. I insult people (R); 17. I sympathize with others' feelings; 22. I am not interested in other people's problems (R); 27. I have a soft heart; 32. I am not really interested in others (R); 37. I take time out for others; 42. I feel others' emotions; 47. I make people feel at ease.

**Neuroticism:** 4. I get stressed out easily; 9. I am relaxed most of the time (R); 14. I worry about things; 19. I rarely feel sad (R); 24. I am easily disturbed; 29. I get upset easily; 34. I change my mood a lot; 39. I have frequent mood swings; 44. I get irritated easily; 49. I often feel sad.

### B HSP-R Scale Items

The Highly Sensitive Person Scale-Revised (HSP-R) comprises 18 items across six subscales, rated on 7-point scale (1=Not at all, 7=Extremely). See [Pluess et al. \(2024\)](#) for complete validation.

**Depth of Processing:** Do you tend to reflect on things deeply?; Do you like deep conversations?; Do you find yourself thinking about philosophical questions?

**Emotional Reactivity:** Are you easily affected by feedback (both negative and positive)?; Do you generally react strongly to your experiences, whether you show it or not?; Are you easily affected by the mood of people around you?

**Overstimulation:** Are you easily overwhelmed by things like bright lights, strong smells, coarse fabrics, or sirens close by?; Do you become unpleasantly aroused when a lot is going on around you?; Are you bothered by intense stimuli, like loud noises or chaotic scenes?

**Sensitivity to Details:** Do you notice when things have been moved around?; Do you tend to notice subtle signs of changing seasons?; Do you seem to notice changes in the weather more than others do?

**Sensitivity to Positive Experiences:** Are you deeply moved by the arts or music?; Do you notice and enjoy delicate or fine scents, tastes, sounds, works of art?; Do you tend to get deeply immersed in music?

**Social Sensitivity:** Do you easily recognize what others are feeling?; Are you good at anticipating how someone may feel about a situation?; Do other people tell you that you are good at understanding what they are feeling or thinking?

### C 50 IPIP-50 Adjectives with Replacements

Complete list of 50 adjectives used in Semantic Factor Analysis, derived from IPIP-50 statements. Four adjectives were unavailable in Word2Vec vocabulary and replaced with semantic equivalents (marked with \*).

**Extraversion (10):** extraverted, *silent\**, unreserved, reserved, talkative, quiet, bold, bashful, assertive, shy

**Agreeableness (10):** selfish, warm, rude, sympathetic, unsympathetic, kind, cold, helpful, considerate, pleasant

**Conscientiousness (10):** organized, disorganized, thorough, careless, prompt, sloppy, neat,

negligent, systematic, careful

**Neuroticism (10):** anxious, relaxed, nervous, imperturbable, touchy, *cranky\**, temperamental, moody, irritable, *sad\**

**Openness (10):** intellectual, unintellectual, imaginative, *incurious\**, innovative, unimaginative, bright, complex, introspective, creative

**Synonym Replacements Used:**

- Item S6 (Extraversion): *untalkative* → *silent*
- Item S29 (Neuroticism): *short-tempered* → *cranky*
- Item S49 (Neuroticism): *self-pitying* → *sad*
- Item S20 (Openness): *uninquisitive* → *incurious*

All replacements were verified to exist in the Google News Word2Vec vocabulary before analysis.

**D HSP-R × Big Five Intercorrelations**

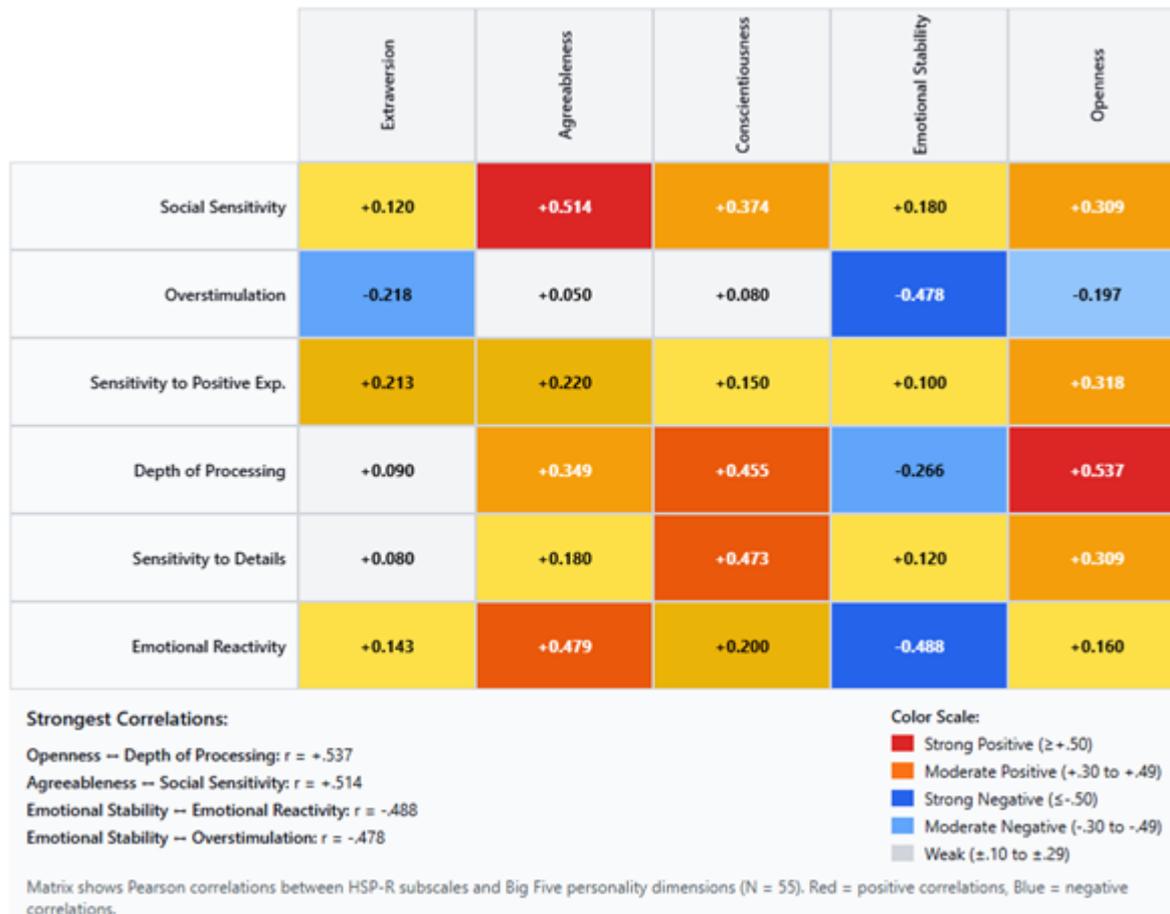


Figure 3: Intercorrelation heatmap between HSP-R subscales and Big Five personality dimensions (N=55)

**E Gender Distributions**

| Dimension                              | Males (%) | Females (%) | Diff. (%) |
|--|-----------|-------------|-----------|
| <i>Big Five Personality Dimensions</i> |           |             |           |
| Emotional Stability                    | 63.1      | 45.4        | -17.7     |
| Extraversion                           | 61.2      | 54.5        | -6.8      |
| Agreeableness                          | 77.4      | 79.8        | +2.3      |
| Conscientiousness                      | 70.6      | 71.0        | +0.4      |
| Openness                               | 77.6      | 75.8        | -1.8      |
| <i>HSP-R Dimensions</i>                |           |             |           |
| Overall HSP-R                          | 66.8      | 75.0        | +8.2      |
| Overstimulation                        | 47.5      | 68.5        | +21.1     |
| Emotional Reactivity                   | 69.4      | 77.4        | +7.9      |
| Sens. to Details                       | 61.7      | 69.6        | +7.9      |
| Sens. to Pos. Exp.                     | 72.8      | 80.6        | +7.8      |
| Depth of Processing                    | 77.7      | 80.8        | +3.1      |
| Social Sensitivity                     | 71.9      | 73.1        | +1.2      |

Table 2: Gender differences Big Five & HSP-R dimensions. All scores in percentages. Negative differences indicate male advantage; positive differences indicate female advantage.