

Personality Traits Predict the Valence but not Semantic Content of Narrative Interpretations

Tory Benson (tory.g.benson@dartmouth.edu)

Department of Psychological & Brain Sciences, Dartmouth College, 3 Maynard St., Hanover, NH 03755 USA

Clara Sava-Segal (csava.gr@dartmouth.edu)

Department of Psychological & Brain Sciences, Dartmouth College, 3 Maynard St., Hanover, NH 03755 USA

Emily S. Finn (emily.s.finn@dartmouth.edu)

Department of Psychological & Brain Sciences, Dartmouth College, 3 Maynard St., Hanover, NH 03755 USA

Abstract

Real-world scenarios often contain ambiguities that lead to differences in interpretation across individuals. These idiosyncrasies may be influenced by one's personality traits. Here, participants viewed ambiguous, naturalistic images and were asked to generate their own interpretation of each. Trait data (positive affect and rumination) were also collected. We used state-of-the-art natural language processing (NLP) models to quantify interpretations' valence and semantic content and examine their relationship to traits. Traits significantly influenced the valence of people's interpretations: higher positive affect predicted more positive valence, while higher rumination predicted more negative valence. In cross-validated analyses, these traits reliably predicted an unseen participant's interpretation valence for a given image above chance. However, interpretations from individuals with similar trait scores were idiosyncratic in their semantic content, indicating that traits predispose valence, but not specific subject matter. Previous studies in this space have been primarily qualitative. Our project underscores how advancements in NLP tools have enabled a more objective, quantitative way to evaluate the role of traits in these interpretive processes.

Keywords: personality; social perception; natural language processing; naturalistic stimuli

Introduction

We are constantly generating interpretations of the world around us, which often differ from those of others. Prior research indicates that these idiosyncrasies in interpretation can be partially attributed to one's intrinsic personality traits (e.g., Finn et al., 2018; Serfass and Sherman, 2013). Specifically, negative trait affect, like depression, has been linked to more negativity in subjective interpretations of one's own life events, while positive trait affect, like well-being, has been linked to more positivity in interpretations (Sherman et al., 2013).

It has long been theorized that one's narrative interpretations of ambiguous images can act as a window into their personalities (e.g., Thematic Apperception Test; Murray, 1943). However, the evaluation process for such tasks often lacks an "objective" truth, relying on qualitative analyses and interrater agreement. In recent years, natural language processing (NLP) tools have streamlined evaluations of such narrative interpretations and allowed for more quantitative measures of their content and valence.

Here, we use a novel paradigm that combines ambiguous, "naturalistic" images with a "MadLibs"-style fill-in-the-blank structure to elicit idiosyncratic, yet highly structured and easily comparable interpretations. We leveraged NLP to quantify features of the narratives—i.e., valence and semantic content—and compare these across subjects to investigate how one's personality traits predict the content of their interpretations.

Methods

Abridged Task Paradigm

Participants ($n = 75$, recruited on Prolific) saw forty-five images, each experimentally validated to elicit diverse interpretations. For each image, participants were tasked with filling out a "MadLibs"-style fill-in-the-blank description with three blanks (see Fig. 1) to generate a narrative describing what they believed was happening. Blanks were created to highlight the more ambiguous parts of each image, such as what the person/people in it are doing or feeling, or where they are located.

Behavioral Measures

Valence Interpretations were assigned positive, negative, and neutral sentiment scores using a RoBERTa-base sentiment language model (Loureiro et al., 2022). Valence was calculated by subtracting the negative sentiment score from the positive score.

Semantic Content To assess similarity in semantic content, each interpretation was converted into vector embeddings using Bidirectional Encoder Representations for Transformers (BERT; Devlin et al., 2019). Due to the fill-in-the-blank structure of the task, while the full sentence was embedded, final embeddings for comparison across subjects were derived only from the content filled into the blanks, allowing the scaffolding sentence to serve as a shared context without artificially inflating similarity values. Cosine similarity was then calculated between these embeddings to quantify the similarity in content between pairs of interpretations.

Semantic Content and Valence are Largely Dissociable Similarity in semantic content are only weakly associated with similarity in valence ($r_s = 0.13$, see Fig. 1 for examples).

Trait Measures

Rumination Rumination was assessed using the 22-item Rumination Response Scale (Treyner et al., 2003). Participants were asked to rate each item (e.g., "Think about a recent situation,

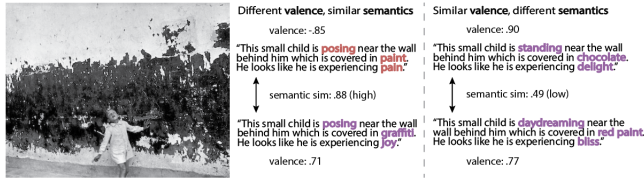


Figure 1: Sample image and interpretations.

wishing it had gone better”) on a Likert scale ranging from 1 (*almost never*) to 4 (*almost always*) based on how often they do each when feeling “down, sad, or depressed.” Scores were approximately normally distributed ($M_{RRS} = 45$, $SD_{RRS} = 14$).

Positive Affect Positive affect was assessed using the Positive and Negative Affect Schedule (Watson et al., 1988). Participants were asked to rate each item (e.g., “Excited”) on a Likert scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*) based on how they feel “on average.” Scores were approximately normally distributed ($M_{PANAS_{pos}} = 31$, $SD_{PANAS_{pos}} = 14$).

Trait Similarity To compare trait scores between participants, we took the Euclidean distance and then inverted these values to get a similarity score.

Statistics

To relate traits to valence, our analyses used linear mixed effects (LME) models with image as a random effect. To relate traits to semantic content, we used an LME model (Chen et al., 2017) to predict semantic similarity from trait similarity with crossed random effects of participant and image as an independent random effect. This approach allowed us to account for non-independence in the data from repeated observations for each participant (pair). To account for redundancy, we manually adjusted the degrees of freedom and standard error (Chen et al., 2017).

Cross-Validation Procedure A 5-fold cross-validation was performed to examine the reliability of our trait-valence effects. We randomly divided participants into 5 folds, training an LME on 4 of the folds and testing on the held-out fold, and repeating this process for all 5 folds. To generate a null distribution to assess prediction accuracy relative to chance, we shuffled trait scores across participants and repeated the cross-validation procedure 1000 times.

Results

Traits Predict Valence but not Semantic Content

Valence Higher rumination scores significantly predict more negative valence used in interpretations across images ($p < .001$, Fig. 2A). Through a 5-fold cross-validation, we find that we can use model weights trained on a subset of our participants to reliably predict valence used given the held-out participants’ rumination scores across individual images ($r = .67$, Fig. 2C).

Higher positive affect scores significantly predict more positive valence used in interpretations across images ($p < .001$, Fig. 2B). Through a 5-fold cross-validation, we find that we can use model

weights trained on a subset of our participants to reliably predict valence used given the held-out participants’ positive affect score across individual images ($r = .67$, Fig. 2D).

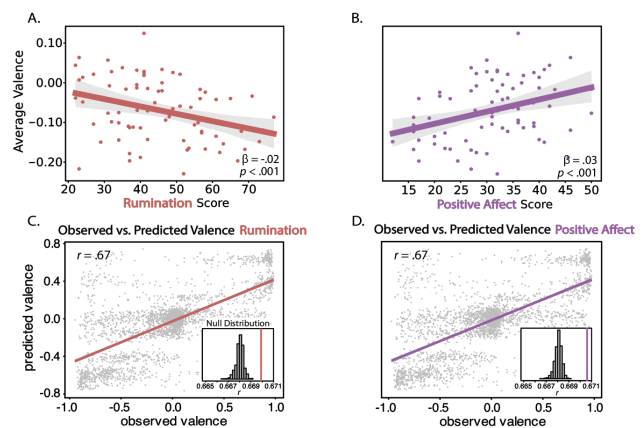


Figure 2: (A) Rumination score predicts more negative valence used (plotted as the average valence score of interpretations across images). (B) Positive affect predicts more positive valence. (C & D) In a 5-fold cross-validation, both rumination and positive affect predict valence beyond image alone.

Taken together, these results suggest that people assign valence to ambiguous, naturalistic images differently depending on their personality traits.

Semantic Content Individuals that are more similar in rumination do not use more similar content in their interpretations ($p = .64$, Fig. 3A). Further, individuals who are more similar in positive affect use less similar content in their interpretations ($p < .001$, Fig. 3B). Taken together, these results suggest that participants with similar traits do not generate interpretations with similar semantic content.

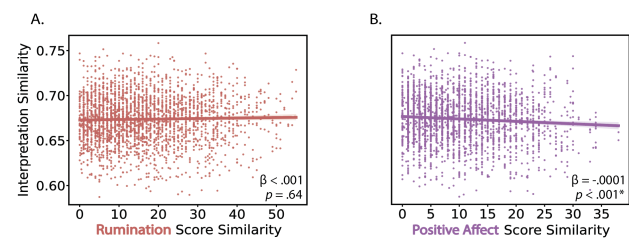


Figure 3: (A) Rumination score similarity does not predict similarity of interpretation content. (B) Similarity in positive affect score predicts less similarity of interpretation content.

Conclusion

Our results suggest that when individuals form subjective interpretations of ambiguous information, their personality traits affect the valence of those interpretations but not their semantic content. Our work also highlights the potential of natural language processing tools to quantify individual differences in narrative interpretations, and how we may use these to understand the influence of people’s intrinsic personality traits on how they see the world.

Acknowledgments

This work was supported by a National Science Foundation Graduate Research Fellowship (to Clara Sava-Segal) and by the National Institutes of Health (grants K99MH120257 and R00MH120257 to Emily S. Finn).

References

- Chen, G., Taylor, P. A., Shin, Y.-W., Reynolds, R. C., and Cox, R. W. (2017). Untangling the relatedness among correlations, Part II: Inter-subject correlation group analysis through linear mixed-effects modeling. *NeuroImage*, 147:825–840.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805 [cs].
- Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A., and Constable, R. T. (2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nature Communications*, 9(1):2043.
- Loureiro, D., Barbieri, F., Neves, L., Anke, L. E., and Camacho-Collados, J. (2022). TimeLMs: Diachronic Language Models from Twitter. arXiv:2202.03829 [cs].
- Murray, H. (1943). *Thematic apperception test*. Harvard University Press.
- Serfass, D. G. and Sherman, R. A. (2013). Personality and perceptions of situations from the Thematic Apperception Test. *Journal of Research in Personality*, 47(6):708–718.
- Sherman, R. A., Nave, C. S., and Funder, D. C. (2013). Situational construal is related to personality and gender. *Journal of Research in Personality*, 47(1):1–14.
- Treynor, W., Gonzalez, R., and Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*, 27(3):247–259.
- Watson, D., Anna, L., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6):1063.