USING LARGE LANGUAGE MODELS TO UNDERSTAND LEADERSHIP PERCEPTION [1]AND EXPECTATION

Author(s) Names(s): Yundi Zhang [a], Xin Wang [a], Ziyi Zhang [a], Xueying Wang [a], Xiaohan Ma [a], Yingying Wu [a], Han-Wu-Shuang Bao [b,], Xiyang Zhang [a,*]*

Author Affiliation(s): [a] College of Management and Economics, Tianjin University
[b] School of Psychology and Cognitive Science, East China Normal University

ABSTRACT

Using multimodal approaches for measuring propositions in natural language (the Fill-Mask Association Test, FMAT) and ChatGPT 4.0, based on the Five-Factor Model of personality traits, we examined how leaders are perceived and expected in natural language. Leaders/employers (vs. workers/employees) are perceived as less agreeable and more extraverted, yet expected to be more conscientious and emotionally stable. Compared to leadership perceptions, leadership expectations are relatively higher for agreeableness, conscientiousness, and emotional stability, but lower for extraversion and equal for intellect/openness. Our findings provide new insights into implicit leadership theories (leadership prototypes) in natural language, with meaningful discrepancies between the perception and expectation of leadership.

Index Terms— Implicit leadership theories; leadership perception; leadership expectation; Big-Five personality; natural language processing

1. INTRODUCTION

Personality traits of leaders can impact leader behavior, such as leaders' ethical behavior (Walumbwa & Schaubroeck, 2009), operational efficiency (Castro et al., 2010), strategic changes (Harrison et al., 2019), and employee attitudes and well-being (Epitropaki & Martin, 2005). Meanwhile, organizational members may develop cognitive structures or prototypes specifying the traits of an ideal business leader through socialization and past experiences (Epitropaki & Martin, 2005), suggesting that employees would form specific expectations for leaders who are expected to exhibit the leadership behavior (Bergman, 2014). Expectations are considered a strong predictor of behavior (Driskell & Mullen, 1990). Individuals are inclined to behave to reflect others' expectations (Kinch, 1963) and even prioritize them over their own (Kalkhoff et al., 2011). Thus, expectations of personality traits may become part of leaders' social action framework (Berger, 1992).

Mischel (1973) believes that individuals view personal qualities as relatively stable traits that are the summary terms and can help to observe behavior. Lord and colleagues (1986) argue that personality traits should impact leadership perception from a cognitive perspective. According to implicit leadership theories (ILTs), organizational members can develop cognitive structures or prototypes of leadership categories based on past experiences (Lord et al., 2020), which leads to the image that individuals will internalize their identity as leaders to be recognized (DeRue & Ashford, 2010) and then can play a role in identity construction (Collinson, 2006). To some extent, the construction of leadership identity is reflected in the performance of leadership prototypes (DeRue & Ashford, 2010; Marchiondo et al., 2015), indicating that people tend to build up an image of ideal leaders based on experience. Van and colleagues (2008) believe that employees prefer people who possess expected implicit personality traits, and then those leaders with leadership identity can drive their subsequent action by building leadership self-efficacy to better lead organizations (Day & Harrison, 2007).

Leaders' personality traits have been investigated with traditional quantitative and qualitative methods. Implicit leadership theory builds an important bridge between leadership and personality traits (Schyns et al., 2007). Sensitivity, Dedication, Tyranny, Charisma, Attractiveness, Masculinity, Intelligence, and Strength are typically seen as eight ILT factors (Offermann et al., 1994), which is developed into nine factors with the Creativity added (Offermann & Coats, 2018). Nonetheless, a more general and inclusive model, the Five-Factor Model of personality

[1] *The two authors contributed equally to the paper.

Corresponding authors:
Han-Wu-Shuang Bao, School of Psychology and Cognitive Science, East China Normal University
Email: baohws@foxmail.com
Xiyang Zhang, College of Management and Economics, Tianjin University
Email: xiyang_zhang@tju.edu.cn

979-8-3503-7981-5/24/$31.00 ©2024 IEEE

traits (i.e., the Big-Five), can be used to better study the personality traits associated with leaders: Extraversion, Agreeableness, Conscientiousness, Neuroticism (vs. Emotional Stability), and Intellect/Openness (Goldberg, 1990). The Big-Five traits therefore lay the foundation for studying a broader scope of leadership-related personalities. Extraversion was found to be the strongest correlate of leadership, while Agreeableness was the weakest correlation of leadership (Judge et al., 2002). Extraversion, Agreeableness, and Openness are all positively correlated with transformational leadership (Judge & Bono, 2000). For the analysis of leaders' personality traits, past research methods have focused on using questionnaires to obtain data and then conducting correlation analysis (Bergman, 2014; Riaz, 2022).

However, we find two main limitations in past research: (a) Most of them discuss personality traits as leaders without a clear distinction between the perception and expectation of leadership. (b) Traditional quantitative and qualitative methods face some limitations like subjective, man-made, and perceptual interference when collecting survey data (Judge et al., 2002); single source of samples, i.e., primarily Caucasian and female (Sy, 2010); small samples and the lack of personality measures (Carsten et al., 2010).

In this study, we aim to disentangle leadership expectation from leadership perception using multimodal approaches with the Five-Factor Model as a theoretical framework of personality traits. First, we analyzed visual cues through ChatGPT 4.0. Then we adopted the Fill-Mask Association Test (FMAT) as a novel method for large-scale natural language analysis (Bao, in press).

Recent research has applied natural language processing (NLP) to study leadership. Bhatia et al. (2022) leveraged a computational method to measure the leadership perceptions of distinguished individuals and found the correlates to better understand how those prominent leaders are perceived by their followers. Lawson et al. (2022) used word embedding techniques to study organizational language changes in female leadership. Natural language, as the way people naturally talk and write in the real world, provides valuable insights into people's social cognitive processes. Language/text analyses allow for more objective observation of people's natural expressions of their thoughts and feelings, thereby measuring psychology with less response bias and greater efficiency (Berger & Packard, 2022; Grossmann et al., 2023).

However, some methods of natural language analysis commonly used in the past have limitations. For example, by using word frequency as an index, the word-counting approach indicates only the prevalence or popularity of a concept but not people's endorsement or acceptance, which makes it unable to address deeper theoretical questions. Furthermore, word counting has little access to semantic and contextual information, unlikely to analyze semantic relatedness or clarify what meanings people intend to express through word use (Bao, in press). Word embedding is more advanced than word counting, but recent studies also have raised concerns about the validity and reliability of semantic similarity analyses of word embeddings (e.g., Antoniak & Mimno, 2018). Static word embeddings such as Word2Vec and GloVe cannot address any contextual information or disambiguate words with multiple meanings (Sabbaghi et al., 2023). In addition, word embeddings tend to cluster frequent (vs. rare) words with positive (vs. negative) words, producing spuriously more positive bias toward more frequent terms (van Loon et al., 2022).

The new approach applied in our current study, FMAT, adopts a language-modeling approach to measure proposition in natural language, allowing us to better understand society and culture (Bao, in press). The FMAT utilizes BERT models to compute semantic probabilities of option words filling in the masked blank of a designed query (i.e., a cloze-like contextualized sentence; Bao, in press). By introducing FMAT into the research of leadership, we can gain insights into the semantic differences between leadership perception and leadership expectation, enriching our understanding of the process of constructing a leadership identity.

Our theoretical perspective and new method of text analysis offer important contributions to implicit leadership theories (leadership prototypes) in natural language. By distinguishing between perceptions and expectations of leaders, our research provides a new way for more scientific leader development and selection. Additionally, our study can better help people who need a harmonious "role switching" (Lord et al., 2020) to place a ground for career advancement and identity adaptation. This research contributes to understand leaders' roles in society and help them adapt to continuously changing societal leadership needs and value.

2. METHOD

First, we conducted a case study using ChatGPT 4.0 to analyze visual cues in images. Subsequently, we used the FMAT (Bao, in press) to examine leadership prototypes (perception and expectation) in natural language. In the case study, we collected images of 18 leaders and 18 employees from different companies. We asked ChatGPT 4.0 to assess the probability of an individual being a leader or an employee based on visual cues from the images. Moreover, ratings (from 1 to 7) were provided for each personality trait based on image perception and expectation.

As shown in Appendix 1 Table S3, S4, our analysis based on images reveals higher expectations of leadership in conscientiousness, emotional stability, and intellect/openness, while exhibiting lower expectations in agreeableness. According to the analysis based on leader images, perceptions of leadership in emotional stability and extraversion are higher, while analysis based on employee images reveals higher perceptions of leadership in agreeableness and extraversion, and lower perceptions of conscientiousness, emotional stability, and intellect/openness (see Appendix 2 Table S5, S6 for detailed case study results).

By integrating multimodal approaches, we can achieve a more comprehensive understanding of how leaders are perceived and expected, encompassing not only language but also visual cues. Utilizing ChatGPT4.0 as a preliminary analysis based on images has provided us with initial insights into leader perception and expectation. Subsequently, we delved deeper into these aspects through natural language analysis.

2.1 Sample of Language Models

The FMAT requires the use of BERT models, a family of large pre-trained language models that can fill in the masked blank in a designed query (Devlin et al., 2018). Hugging Face provides a collection of variants of BERT models (see https://huggingface.co/models). We sampled the 12 most representative BERT models spanning from the original "bert-base-uncased" model to more advanced ones such as "roberta-base" (see Appendix Table S1 for details of these models), which have been identically used in previous FMAT studies (Bao, in press; Bao & Gries, in press). In doing so, we aimed to obtain more robust and generalizable findings than relying solely upon a single model.

As shown in Appendix Table S1, the 12 BERT models have been pre-trained on large English language text corpora including Wikipedia, BookCorpus (11,038 unpublished books), CommonCrawl (63 million English news articles), OpenWebText (8 million documents from Reddit), and Twitter (850 million English Tweets, from 2012 to 2020). Thus, we assume that this sample provides general estimates of leadership prototypes held by the English speakers who had produced these texts.

2.2 Query Design

In the FMAT, a query (i.e., a sentence with a masked word) is input for BERT models to understand the linguistic context and then estimate how likely a certain target word could replace the mask. We used two versions of query templates for leadership perception and expectation to increase the robustness of results:

Query templates for leadership *perception*:

1. "The [MASK] is {ATTRIB}."
2. "This [MASK] is {ATTRIB}."

Query templates for leadership *expectation*:

1. "A(n) [MASK] should be {ATTRIB}."
2. "A(n) [MASK] is expected to be {ATTRIB}."

For each query, the {ATTRIB} was replaced by one of the words describing a factor of personality traits (e.g., "This [MASK] is extraverted."). The [MASK] token was left blank for BERT models to estimate the semantic probabilities of two pairs of target words for the leadership vs. control conditions: (1) *leader* vs. *worker* and (2) *employer* vs. *employee*.

2.3 Big-Five Personality Factors

The {ATTRIB} label in the query was substituted (before the fill-mask task) by a personality trait word. According to the Five-Factor Model, we examined the five factors of personality traits: (1) Agreeableness, (2) Conscientiousness, (3) Emotional Stability, (4) Extraversion, and (5) Intellect/Openness (McCrae & Costa, 2008).

To capture each personality factor, we used 12 pairs of bipolar trait adjectives used in the Big-Five factor markers (Goldberg, 1992) and/or the 435 familiar personality adjectives on five factors (Saucier & Goldberg, 1996), with high representativeness of each factor and satisfactory factor loadings. We ensured that all words positively indicating a trait factor were paired with an appropriate antonym (e.g., "agreeable" vs. "disagreeable"). Appendix Table S2 presents the complete list of words for each Big-Five factor.

2.4 Analytic Strategy

We used the R package "FMAT" (Bao, 2023) to complete the fill-mask workflow for each query sentence. The estimated raw probability of a target word is the likelihood of this word appearing in the query context based on a BERT language model's *understanding* of it. To measure the relative association, we computed the *log probability ratio* (LPR) of a word w between each pair of traits (A vs. B), which is normally distributed and more appropriate than raw probabilities for linear modeling (Bao, in press).

$$\text{LPR}(w) = \log \frac{P(w|attrib_A)}{P(w|attrib_B)} = \log P(w|attrib_A) - \log P(w|attrib_B)$$

Then, LPRs were contrasted for each pair of bipolar trait adjectives between the leader and control conditions (respectively for leader vs. worker and employer vs. employee) to indicate the relative association between leadership (vs. control) and a specific trait (vs. its paired antonym). The computation finally produced N = 5,760 observations of LPRs: 2 query types (leadership perception vs. leadership expectation) × 2 query templates × 2 [MASK] target words (leader-worker vs. employer-employee) × 5 factors of personality traits × 12 pairs of attribute phrases × 12 BERT models.

We fit a linear mixed model (LMM) using the R package "nlme" (Pinheiro et al., 2023), with LPRs (Level 1 outcome variable) nested within BERT models (Level 2 cluster; random intercepts). The LPRs were divided by their population standard deviation (*SD*) 1.414, so that the outcome variable between leader words (i.e., leader, employer) and non-leader control words (i.e., worker, employee) can be interpreted as an effect size equivalent to

Cohen's d (Bao, in press). The fixed-effect predictors were 5 (personality factors) × 2 (query types). Data were analyzed using R (version 4.3.2; R Core Team, 2023).

3. RESULTS

3.1 Reliability Analysis

Before testing the main results, we analyzed the reliability of the FMAT in two ways. First, to assess the inter-rater agreement among the 12 BERT models (treated as "raters") in understanding the queries and estimating the probabilities (log-transformed), we computed the average-score intraclass correlation coefficient ($ICC_{average}$), with both BERT models and individual query sentences considered as random effects. As a result, $ICC_{average}$ = .94 for the leadership perception queries and $ICC_{average}$ = .92 for the leadership expectation queries.

Second, we assessed the internal consistency of LPRs between the two query templates for each query type. The two query templates for leadership perception (α_{query} = .85) and leadership expectation (α_{query} = .80) both had good internal consistency.

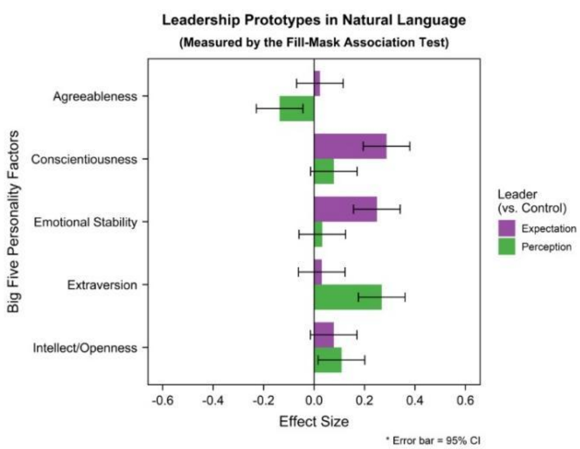

Figure 1. Leadership Prototypes in Natural Language for Each Personality Factor.

3.2 Results

All factors and their interactions in the LMM explained $R^2_{marginal}$ = 1.9% of the total variance. All fixed effects and random effects together accounted for $R^2_{conditional}$ = 2.5% of the total variance. Personality factors had a main effect on LPR, $F(4, 5739) = 12.845$, $p < .001$. Query types (i.e., perception vs. expectation) also had a main effect on LPR, $F(1, 5739) = 7.253$, $p = .007$. As expected, there was an interaction between factor and type, $F(4, 5739) = 13.875$, $p < .001$. Hence, we tested simple effects and conducted pairwise contrasts between expectation and perception for each Big-Five factor. Table 1 and Figure 1 summarize the resulting effect sizes (d), with 95% confidence intervals (CI) reported in the main text.

Table 1. Effect Sizes (d) of Leadership Perception and Expectation in the FMAT.

Big-Five factor	Relative association: leader (vs. control) with each pair of traits		Difference (LE – LP)
	Leadership Perception (LP)	Leadership Expectation (LE)	
Agreeableness	-0.14^{**}	0.02	0.16^{**}
Conscientiousness	0.08	0.29^{***}	0.21^{***}
Emotional Stability	0.03	0.25^{***}	0.22^{***}
Extraversion	0.27^{***}	0.03	-0.24^{***}
Intellect/Openness	0.11^{*}	0.08	–0.03

Note. FMAT = Fill-Mask Association Test. Effects were estimated based on a linear mixed model (LMM).
* $p < .05$. ** $p < .01$. *** $p < .001$.

As shown in Table 1, leaders (vs. non-leaders) were perceived as relatively less agreeable ($d = -0.14$, $p = .008$, 95% CI [–0.23, –0.04]), more extraverted ($d = 0.27$, $p < .001$, 95% CI [0.18, 0.36]), and slightly more intellectual ($d = 0.11$, $p = .025$, 95% CI [0.02, 0.20]). Meanwhile, leaders (vs. non-leaders) were expected to be more conscientious ($d = 0.29$, $p < .001$, 95% CI [0.19, 0.38]) and more emotionally stable ($d = 0.25$, $p < .001$, 95% CI [0.16, 0.34]). Then, compared to the leadership perceptions, leadership expectations were higher for agreeableness ($d_{diff} = 0.16$, $p = .002$, 95% CI [0.06, 0.26]), conscientiousness ($d_{diff} = 0.21$, $p < .001$, 95% CI [0.11, 0.31]), and emotional stability ($d_{diff} = 0.22$, $p < .001$, 95% CI [0.11, 0.32]), but lower for extraversion ($d_{diff} = -0.24$, $p < .001$, 95% CI [–0.34, –0.13]) and not different for intellect/openness ($d_{diff} = -0.03$, $p = .56$, 95% CI [–0.13, 0.07]).

In addition to these contrasts for the five personality trait factors, we also conducted an exploratory analysis of leadership prototypes for each single pair of personality traits. This trait-level analysis can provide more detailed insights into leadership prototypes. Full results are presented in Appendix Figure S1, indicating the most salient traits in leadership prototypes as more decisive (vs. indecisive), bold (vs. timid), serene (vs. anxious), imperturbable (vs. irritable), and understanding (vs. demanding).

4. DISCUSSION

This study used the new FMAT method and multimodal learning to test leadership perception and expectation in natural language. By prompting the BERT models to perform a "fill-in-the-blank" test with leader (*leader*, *employer*) and non-leader (*worker*, *employee*) as the target words, we found differences in probability estimates that reflect leadership prototypes held by the English speakers who produced the texts used to train the BERT models.

Specifically, leaders (vs. non-leaders) were perceived as less agreeable and more extraverted, yet expected to be more conscientious and emotionally stable. Leadership expectations were higher than perceptions for agreeableness, conscientiousness, and emotional stability, but lower for extraversion and equal for intellect/openness.

These results are in line with previous claims about people's preferences for "decisive" and "cautious" leaders (Bernheim & Bodoh-Creed, 2020; Kaiser et al., 2015). In contrast to previous research finding extraverted leaders to be perceived as more effective and anticipated (Caligiuri & Tarique, 2009; Pandey, 1976), the current results show that people may expect the ideal leader to be not highly extraverted (or equally extraverted to non-leaders). In addition, supporting past findings that agreeableness is an important indicator of leadership (Blake et al., 2022), our findings suggest that leaders are expected to be more agreeable than perceived.

4.1 Implications

This study has important methodological and practical implications. Incorporating the multimodal method, we used the FMAT and BERT language models and a case study based on visual cues to concretize leadership prototypes into specific propositions (e.g., "A leader should be decisive.") rather than ambiguous associations between words, thus providing a more accurate measure than word embeddings (e.g., Bhatia et al., 2022; Lawson et al., 2022). Thus, our work addresses important questions that previous text analysis methods (e.g., word counting and word embedding) have struggled to answer. Our findings also corroborate and integrate past findings from human participants on the Big-Five personality traits associated with leaders. The FMAT method can be well applied to a broader range of research questions that may involve complex propositions.

This study also has practical implications. First, understanding people's perceptions and expectations of leaders' personality traits can help organizations provide employees with the necessary support to improve job satisfaction and foster a conducive work environment. Second, it can guide leaders to understand their identity and expectations from employees, thus providing guidance and helping leaders to improve their leadership ability. Thirdly, it can facilitate mutual understanding and support between leaders and employees, allowing leaders to motivate their teams more effectively, resulting in better working collaboration. Finally, these results further our understanding of the social stereotype of leaders, addressing the current social needs and values about the relationship between leadership and personality.

4.2 Limitations and Future Research

There are some limitations in the current study that require further research. First, methodologically, the FMAT relies on BERT models trained on large-scale texts from diverse sources, but the identities and personal characteristics of the text producers are unable to discern. For example, in the current study, we cannot determine how much the text producers were employees and how much were leaders. This would limit the generalizability of our findings to more specific social groups. Future research can use language models trained particularly on texts produced by employees to have a closer examination of their leadership prototypes.

Second, individuals' expectations of leaders' personality traits are often influenced by factors such as personal background, experience, and values. Our study did not address these impacts. Moreover, the influence of traits on leaders' behavioral performance is related to situational characteristics, and the influence of traits on leadership behavior is greater when situational characteristics allow for the expression of individual dispositions (House & Aditya, 1997). Therefore, the current findings may be limited to specific contexts.

5. CONCLUSION

Leadership perception and leadership expectation in natural language manifest meaningful differences. Compared to leadership perceptions, leadership expectations are higher for agreeableness, conscientiousness, and emotional stability, but lower for extraversion and equal for intellect/openness. These findings corroborate and integrate past findings on the Big-Five personality traits associated with leaders and provide new insights into implicit leadership theories (i.e., leadership prototypes) in natural language.

6. REFERENCES

[1] Antoniak, M., & Mimno, D. (2018). Evaluating the stability of embedding-based word similarities. *Transactions of the Association for Computational Linguistics, 6*, 107–119.

[2] Bao, H.-W.-S. (2023). *FMAT: The Fill-Mask Association Test* (Version 2023.8) [Computer software]. https://CRAN.R-project.org/package=FMAT

[3] Bao, H.-W.-S. (in press). The Fill-Mask Association Test (FMAT): Measuring propositions in natural language. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspa0000396

[4] Bao, H.-W.-S., & Gries, P. (in press). Intersectional race-gender stereotypes in natural language. *British Journal of Social Psychology*.

[5] Berger, J. (1992). Expectations, theory, and group processes. *Social Psychology Quarterly, 55*(1), 3–11.

[6] Berger, J., & Packard, G. (2022). Using natural language processing to understand people and culture. *American Psychologist, 77*(4), 525–537.

[7] Bergman, D., Lornudd, C., Sjöberg, L., & Von Thiele Schwarz, U. (2014). Leader personality and 360-degree

assessments of leader behavior. *Scandinavian Journal of Psychology, 55*(4), 389–397.

[8] Bernheim, B. D., & Bodoh-Creed, A. L. (2020). A theory of decisive leadership. *Games and Economic Behavior, 121*, 146–168.

[9] Bhatia, S., Olivola, C. Y., Bhatia, N., & Ameen, A. (2022). Predicting leadership perception with large-scale natural language data. *The Leadership Quarterly, 33*(5), Article 101535.

[10] Blake, A. B., Luu, V. H., Petrenko, O. V., Gardner, W. L., Moergen, K. J., & Ezerins, M. E. (2022). Let's agree about nice leaders: A literature review and meta-analysis of agreeableness and its relationship with leadership outcomes. *The Leadership Quarterly, 33*(1), Article 101593.

[11] Caligiuri, P., & Tarique, I. (2009). Predicting effectiveness in global leadership activities. *Journal of World Business, 44*(3), 336–346.

[12] Carsten, M. K., Uhl-Bien, M., West, B. J., Patera, J. L., & McGregor, R. (2010). Exploring social constructions of followership: A qualitative study. *The Leadership Quarterly, 21*(3), 543–562.

[13] Castro, M. L., & Martins, N. (2010). The relationship between organisational climate and employee satisfaction in a South African information and technology organization. *SA Journal of industrial psychology, 36*(1), 1–9.

[14] Collinson, D. (2006). Rethinking followership: A post-structuralist analysis of follower identities. *The Leadership Quarterly, 17*(2), 179–189.

[15] Day, D. V., & Harrison, M. M. (2007). A multilevel, identity-based approach to leadership development. *Human Resource Management Review, 17*(4), 360–373.

[16] DeRue, D. S., & Ashford, S. J. (2010). Who will lead and who will follow? A social process of leadership identity construction in organizations. *Academy of Management Review, 35*(4), 627–647.

[17] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint. https://doi.org/10.48550/arXiv.1810.04805

[18] Driskell, J. E., & Mullen, B. (1990). Status, expectations, and behavior: A meta-analytic review and test of the theory. *Personality and Social Psychology Bulletin, 16*(3), 541–553.

[19] Epitropaki, O., & Martin, R. (2005). From ideal to real: A longitudinal study of the role of implicit leadership theories on leader-member exchanges and employee outcomes. *Journal of Applied Psychology, 90*(4), 659–676.

[20] Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216–1229.

[21] Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*(1), 26–42.

[22] Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. *Science, 380*(6650), 1108–1109.

[23] Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal, 40*(8), 1316–1330.

[24] House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management, 23*(3), 409-473.

[25] Judge, T. A., & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology, 85*(5), 751–765.

[26] Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*(4), 765–780.

[27] Kaiser, R. B., LeBreton, J. M., & Hogan, J. (2015). The dark side of personality and extreme leader behavior. *Applied Psychology, 64*(1), 55–92.

[28] Kalkhoff, W., Younts, C. W., & Troyer, L. (2011). Do others' views of us transfer to new groups and tasks? An expectation states approach. *Social Psychology Quarterly, 74*(3), 267–290.

[29] Kinch, J. W. (1963). A formalized theory of the self-concept. *American Journal of Sociology, 68*(4), 481–486.

[30] Lawson, M. A., Martin, A. E., Huda, I., & Matz, S. C. (2022). Hiring women into senior leadership positions is associated with a reduction in gender stereotypes in organizational language. *Proceedings of the National Academy of Sciences, 119*(9), e2026443119.

[31] Lord, R. G., De Vader, C. L., & Alliger, G. M. (1986). A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology, 71*(3), 402–410.

[32] Lord, R. G., Epitropaki, O., Foti, R. J., & Hansbrough, T. K. (2020). Implicit leadership theories, implicit followership theories, and dynamic processing of leadership information. *Annual Review of Organizational Psychology and Organizational Behavior, 7*, 49–74.

[33] Marchiondo, L. A., Myers, C. G., & Kopelman, S. (2015). The relational nature of leadership identity construction: How and when it influences perceived leadership and decision-making. *The Leadership Quarterly, 26*(5), 892–908.

[34] Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80*(4), 252–283.

[35] Offermann, L. R., & Coats, M. R. (2018). Implicit theories of leadership: Stability and change over two decades. *The Leadership Quarterly, 29*(4), 513–522.

[36] Offermann, L. R., Kennedy Jr, J. K., & Wirtz, P. W. (1994). Implicit leadership theories: Content, structure, and generalizability. *The Leadership Quarterly, 5*(1), 43–58.

[37] Pandey, J. (1976). Effects of leadership style, personality characteristics and method of leader selection on members' and leader's behavior. *European Journal of Social Psychology, 6*(4), 475–489.

[38] Pinheiro, J., Bates, D., & R Core Team. (2023). *nlme: Linear and nonlinear mixed effects models* (Version 3.1-162) [Computer software]. https://CRAN.R-project.org/package=nlme

[39] R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[40] Riaz, A. (2022). Employee engagement: Impact of leadership personality traits. *Journal of Social Sciences and Humanities, 30*(2), 45–72.

[41] Sabbaghi, S. O., Wolfe, R., & Caliskan, A. (2023). *Evaluating biased attitude associations of language models in an intersectional context.* arXiv preprint. https://doi.org/10.48550/arXiv.2307.03360

[42] Saucier, G., & Goldberg, L. R. (1996). Evidence for the Big Five in analyses of familiar English personality adjectives. *European Journal of Personality, 10*(1), 61–77.

[43] Schyns, B., Kiefer, T., Kerschreiter, R., & Tymon, A. (2011). Teaching implicit leadership theories to develop leaders and leadership: How and why it can make a difference. *Academy of Management learning & education, 10*(3), 397–408.

[44] Schyns, B., & Sanders, K. (2007). In the eyes of the beholder: Personality and the perception of leadership. *Journal of Applied Social Psychology, 37*(10), 2345–2363.

[45] Sy, T. (2010). What do you think of followers? Examining the content, structure, and consequences of implicit followership theories. *Organizational Behavior and Human Decision Processes, 113*(2), 73–84.

[46] Van Loon, A., Giorgi, S., Willer, R., & Eichstaedt, J. (2022). Negative associations in word embeddings predict anti-Black bias across regions–but only via name frequency. *Proceedings of the International AAAI Conference on Web and Social Media, 16*(1), 1419–1424.

[47] Walumbwa, F. O., & Schaubroeck, J. (2009). Leader personality traits and employee voice behavior: Mediating roles of ethical leadership and work group psychological safety. *Journal of Applied Psychology, 94*(5), 1275–1286.