scientific reports

OPEN

Check for updates

Cultural differences in the beauty premium

Benjamin Kohler¹ & Wladislaw Mill²⊠

A large body of research suggests that better-looking people are associated with various positive outcomes, an effect labeled the beauty premium. However, the vast majority of this evidence has been demonstrated in WEIRD (Western, Educated, Industrialized, Rich, and Democratic) countries, neglecting the role of cultural mechanisms in translating beauty into socio-economic outcomes. We close this gap by leveraging cultural differences embedded in an essential artifact of culture: language. Specifically, we establish a method for examining beauty associations in machine learningbased language models. Using this method, we replicate several standard findings in English. More importantly, we create a linguistic measure of the beauty premium and apply it to 68 languages. We provide new evidence that the beauty premium may be universal, with considerable heterogeneity across cultures. Furthermore, we discuss several new avenues for future research arising from our findings.

Keywords Beauty premium, Language models, Cultural differences

Physical appearance is a crucial characteristic with social and economic benefits to the mere attractiveness of individuals, widely known as the beauty premium¹⁻⁴. While some research indicates that attractive people may face "beauty penalties," and unattractive individuals may even receive benefits from "ugliness", most research suggest that more attractive individuals are generally evaluated and treated more positively⁶, are more likely to be hired⁷, persistently receive higher wages over their lifetimes⁸, receive better teaching quality ratings in academia², and are even more likely to be elected in democratic elections³.

The explanations for the beauty premium are wide-ranging, from greater confidence affecting bargaining behavior⁹, stereotypical positive associations between beauty and positive personality traits¹⁰, and ascribed intellectual ability¹¹, to evolutionary explanations stipulating that the beauty premium results from positive physical appearance signaling better genes and mate selection¹²⁻¹⁴. While most of these explanations suggest a universal beauty premium, research on the beauty premium has been conducted almost exclusively in WEIRD (Western, Educated, Industrialized, Rich, and Democratic) countries¹⁵, with only few studies focusing on the beauty premium in a non-WEIRD setting (i.e., China)^{16,17}.

Although culture is highly predictive of social and economic behavior¹⁸⁻²¹, differences in the beauty premium across cultures have not been systematically examined. Culturally upheld values²², norms²³, religious beliefs²⁴, and narratives²⁵ all serve as explanatory factors for the differences of market institutions and outcomes observed worldwide. Despite its potential relevance, the role of culture in social and economic effects of physical appearance has been largely overlooked in the literature. Even, cross-cultural comparisons examining beauty stereotypes are scarce and conducted on a small scale^{26,27}. Moreover, these stereotypes have not been linked to the beauty premium, and more generally, there is no study to date investigating the beauty premium across cultures.

Our study fills this gap by examining how the beauty premium differs across cultures. To do so, we rely on a fundamental aspect of culture: language. Specifically, we use machine learning-based language models, called word embeddings²⁸, trained on a vast amount of text data from all over the internet²⁹⁻³¹, which enables us to 1) analyze the association between beauty and success in English and 2) investigate this association across a variety of cultures. Conceptually, word embeddings assign a vector to each word in a vocabulary, capturing the semantic relationships between words by positioning those that are more associated with each other (e.g., circle and ball) closer together in the multidimensional vector space than those that are less related (e.g., circle and cat). Thus, the geometrical similarity between two word vectors can be interpreted as their semantic similarity in the respective language. Word embeddings trained on a vast amount of diverse texts have been used to study specific settings (e.g. emotionality in US Congress' speeches³², and economic training on judges ruling³³), and cultural properties reflected in a language more broadly^{34–37}. For example, it has been documented that implicit biases^{34,38} — particularly gender biases^{36,37} — social class associations³⁵, and bodyweight-related stereotypes³⁹ are all reflected within these word embeddings (see SI A for more detail). Following this literature, we identify

¹Center for Law & Economics, ETH Zurich, Haldeneggsteig 4, 8092 Zurich, Switzerland. ²Department of Economics, University of Mannheim, L7 3-5, 68131 Mannheim, Germany. Zemail: mill@uni-mannheim.de

the beauty premium in word embeddings. First, we extract relevant concepts based on the literature^{34-36,39}, such as beauty, ugliness, and success. To capture the nuanced shades of meaning associated with our concepts of interest, we represent them by a set of words. Building on an existing framework⁴⁰, we aggregate the strength of association between terms representing beauty to terms representing success and failure, effectively constructing a measure for the beauty premium within word embeddings. Finally, we apply our measure to word embeddings pre-trained on a large, diverse set of texts in 68 different languages to explore cross-cultural aspects of the beauty premium. This allows us to investigate the extent to which the beauty premium is ingrained in language usage worldwide.

There are two important findings from our analysis. First, we provide evidence that English word embeddings reveal a clear association between beauty and success. Thus, we establish that the beauty premium is indeed ingrained in the English language, which we call the *linguistic beauty premium*. On the other hand, we also find a stronger association of ugliness with failure, i.e., a linguistic ugliness penalty. Second, and most importantly, we focus on cultural differences in the beauty premium. The cross-language comparison reveals substantial heterogeneity in the existence and size of the linguistic beauty premium across cultures. The linguistic beauty premium ranges from non-existent in some languages (like Burmese and Vietnamese) to estimates up to four times larger than the English estimate (like Finnish and Japanese). However, the vast majority of languages reveal a clear and strong linguistic beauty premium. In summary, our findings suggest that a) the beauty premium is culturally embedded as reflected in language use, and b) this linguistic beauty premium varies across the world but is very persistently found around the globe.

Results

The linguistic beauty premium

First, to validate our approach, we test whether beauty is more associated with something 1) good and 2) pleasant⁶. To do this, we exploit the geometric similarity between word vectors as a measure of semantic proximity or association. We combine the average association of multiple, previously tested words representing *Beauty/Ugliness* (e.g., pretty, lovely, cute, etc.) with a target dimension (e.g., *Good* vs. *Bad*) using the POLAR framework (see Materials and Methods for more details on the POLAR framework and the word selection). The framework yields a single, interpretable association score for *Beauty/Ugliness* on the target dimension. In Fig. 1, we illustrate these association scores with several target dimensions (see SI B for the associations of the individual *Beauty/Ugliness* on the respective dimension (e.g., *Good/Bad*). Positive values mean that *Beauty-* or



Figure 1. English *Beauty/Ugliness* association with other categories. The figure illustrates how *Beauty* and *Ugliness* are associated with a set of traits. Green dots denote the average association of all words representing *Beauty* on a two-poled scale (e.g., *Good - Bad* scale as a continuous measure). Red dots indicate the average association of all words representing *Ugliness* on the same scale. Positive values mean that words of either *Beauty* or *Ugliness* are, on average, more associated with the **right pol** of the scale (e.g., *Good*), while negative values mean that words of either *Beauty* or *Ugliness* are, on average, more associated with the **right pol** of the scale (e.g., *Good*). While negative values mean that words of either *Beauty* or *Ugliness* are, on average, more associated with the **right pol** of the scale (e.g., *Good*). While negative values mean that words of either *Beauty* or *Ugliness* are, on average, more associated with the **right pol** of the scale (e.g., *Bad*). Whiskers denote the 95% confidence interval.

Ugliness-words are, on average, more associated with the right pole of the dimension (e.g., *Good*), while negative values denote that *Beauty*- or *Ugliness*-words are, on average, more associated with the left pole of the dimension (e.g., *Bad*) (see SI C for regressions of the difference between *Beauty* and *Ugly* with regard to these dimensions). *Beauty*-words are, on average, much more associated with *Good/Pleasant* than with something *Bad/Unpleasant*. Similarly, *Ugliness*-words are, on average, much more associated with *Bad/Unpleasant* than with *Good/Pleasant*. The difference between *Beauty* and *Ugliness* on both dimensions is highly significant ($p \le 0.001$). When focusing on all individual words (see SI B), we find that essentially all *Beauty*-words are closer, in terms of language space, to *Good* and *Pleasant* than to *Bad* and *Unpleasant*, with mirrored observations for all *Ugliness*-words. Thus, we find evidence that our construct of *Beauty/Ugliness* indeed reflects intuitive associations and is also consistent with the 'what-is-beautiful-is-good' stereotype, whereby more attractive individuals are generally viewed more positively on a variety of traits^{10,41}.

To obtain a measure of our main objective — the beauty premium — we estimate the association of *Beauty* on the *Success - Failure* dimension and, additionally, on the *Affluence - Poverty* dimension. *Beauty* is substantially more associated with *Success* and *Affluence* than with *Failure* and *Poverty*. *Ugliness*, on the other hand, is clearly associated with *Failure* and *Poverty*, revealing the presence of a linguistic ugliness penalty. These differences are again highly significant for both dimensions. Zooming in on the individual words representing *Beauty* and *Ugliness*, we find that the vast majority of the words resemble the same pattern. Therefore, we can conclude the existence of a beauty premium in the English language, highlighting that the association of beauty and success is deeply rooted in culture.

A major advantage of using language to study the beauty premium is that we can directly examine whether beauty is associated with common explanations for the beauty premium, specifically perceived competence^{4,9,11}, trust^{42–44}, and confidence^{6,9}. For all three dimensions, we find the associations established in the literature. *Beauty* is more associated with *Competence*, *Trust*, and *Confidence* while *Ugliness* is associated with *Incompetence*, *Distrust*, and *Doubt*. Taken together, this evidence replicates — using the English language — some of the most common explanations for the beauty premium.

In SI D, we further validate that these results are specific to *Beauty/Ugliness*, by exploring two unrelated dimensions: *Fast/Slow* and *Far/Near*. We document that the *Fast/Slow* dimension does not exhibit the pronounced associations observed with *Beauty/Ugliness*, and that neither *Far*-words nor *Near*-words show meaningful associations with the relevant categories. These findings further support the conclusion that the linguistic beauty premium is specific to beauty. However, it remains possible that other words related to *Beauty/Ugliness* may exhibit similar patterns, as suggested by approaches in prior research see^{45,46}.

Cultural differences in the beauty premium

Next, we compare the linguistic beauty premium across cultures by comparing the association of *Beauty* between *Success* and *Failure* in each language word embedding. Figure 2 reports the linguistic beauty premium of all the 68 languages we use (see SI E for an estimation using a hierarchical linear mixed-effects model). The vast majority of languages (63 out of 68) show a positive and mostly significant association of beauty with success, suggesting that the beauty premium may be universal. Strikingly, we find substantial heterogeneity. For example, Hindi and French both have a stronger linguistic beauty premium than English. At the same time, some languages, like Burmese and Vietnamese, have even a negative association between *Beauty* and *Success*. From this, we can derive three main insights: (1) most languages show a linguistic beauty premium, (2) English is relatively mediocre in the extent of its beauty premium, and (3) some languages have no or even a negative linguistic beauty premium.

To study how the beauty premium is distributed across the world, we link the 68 languages in our sample to all countries using them as their primary language, illustrated in Fig. 3. With the exception of very few places (like Vietnam and Romania), we find a positive association between *Beauty* and *Success* across the world. Essentially, every continent makes this association (in SI F, we further zoom into Europe as we observe the biggest variation there). This result suggests an almost universal association of *Beauty* and *Success*, revealing a global beauty premium.

Discussion and conclusion

In this study, we ask whether the beauty premium is reflected in language use and to what extent this culturally encoded beauty premium persists across the globe. Based on the POLAR framework⁴⁰, we derive a meaningful representation of beauty and ugliness in word embeddings trained on large text corpora from various sources, allowing conclusions about general language use^{34,35,47,48}. We demonstrate that this measure reflects basic stereotypes such as the association of beauty and pleasantness⁶. More importantly, we uncover a linguistic beauty premium and ugliness penalty in English word embeddings. We also show that biased beliefs about beauty, specifically greater trustworthiness, competence, and self-confidence, are also reflected in word embeddings. Finally, we find that the linguistic beauty premium is present in most of the languages studied but that the magnitude of the linguistic beauty premium shows considerable variance across languages.

By showing that the beauty premium is embedded in broader language use in many societies around the world, we highlight the importance of culture as a little-considered mediating channel for the beauty premium. Understanding the beauty premium as a culturally learned bias implies that the beauty premium may change as cultural norms, values, and narratives change, and attempts to mitigate the beauty premium must consider idiosyncratic cultural associations.

Our results provide strong support for — and a new perspective on — previously found biased beliefs as explanations for the beauty premium. The fact that these biased beliefs are reflected in language may explain why the beauty premium vanishes in experimental labor market settings when employers are informed about actual work performance⁴, but the beauty premium is still found in many real-world markets. While employers may temporarily update their beliefs, the cultural bias could serve as a default to which employers return when either



Figure 2. Linguistic beauty premium by language. The figure illustrates how *Beauty* is associated with *Success* relative to *Failure* (i.e., the linguistic beauty premium) across the 68 languages. Languages displayed in green denote a significantly positive linguistic beauty premium in the respective language. Languages displayed in gray and red denote a non-significant and a significantly negative linguistic beauty premium in the relative language, respectively. Whiskers denote the 95% confidence interval.

the performance signals are ambiguous or when they face a long time between these signals and the bargaining situation⁴⁹.

Our study joins recent work linking language use to economic outcomes^{50,51}. First, by showing that the beauty premium is encoded in language use, our study highlights the importance of culture for the beauty premium. As language is not only a part of culture in itself but can reflect various cultural aspects such as norms or values, our study of language provides a cross-sectional view of the cultural beauty associations. Consequently, the stability of the beauty premium depends on the stability of the underlying cultural stereotypes, and policies attempting to mitigate the effects of the beauty premium should take into account the particularities of each culture.

Second, our research supports theories explaining the beauty premium as a result of beauty signaling individual traits^{9–11,44}. By showing that associations between beauty and various traits found in experimental settings are replicated in word embeddings, the results suggest that the beauty premium may be rooted in entrenched cultural stereotypes.

Third, our results present new hypotheses for the size of the beauty premium for countries around the world, for example, comparing the size of the beauty premium in Japan and Vietnam. In addition, word embeddings



Figure 3. The linguistic beauty premium across the world. The primary language of each country is taken to establish the linguistic beauty premium. Green, ochar shaded countries indicate a positive and negative linguistic beauty premium, respectively. Grey-shaded countries speak, as their primary language, a language not part of our sample. Figure created with *R version 4.2.2*.

could be used to find more universal and culturally specific associations of beauty. Thus, our approach can serve as a fruitful hypothesis-generating tool to find novel explanations for the beauty premium.

Our study is not without limitations. First, our method depends on the selected categories and words. We address this issue by using a multi-word representation and - where possible - previously established word sets. Further, we present our results for individual words and multiple categories, reducing the risk of selected results. Nevertheless. it might be possible that using different sets of words might lead to slightly different results. Second, our analysis only offers a static view on language. As cultural norms are subject to constant (but slow) change, we cannot rule out that our analysis depicts only a very recent status. Future research could analyze the stability of the linguistic beauty premium over time by training a word embedding enriched with temporal information. Third, by approximating cultural properties with language use, we miss cultural differences within the same language area. This especially limits the interpretability of languages with a widespread use in different cultures (e.g., French) and the detection of regional patterns. Given the lack of pre-trained word embeddings at the regional level and the trade-off between granularity and scope of regional analysis, we deliberately focused on establishing the novel linguistic beauty premium and applying it to a variety of languages spoken around the world. This approach is analogous to other research establishing a linguistic gender bias across languages³⁷. However, we also encourage follow-up research to focus on certain demographic sub-groups, regions, and temporal dynamics of the linguistic beauty premium. Fourth, we study the cultural beauty premium using one specific - but highly relevant - factor of a culture: the language. However, we do not know whether and how our results would translate to other markers of culture. Future research might want to extend our approach to either other markers of culture or conduct other cross-cultural investigations into the beauty premium. Fifth, the evidence we provide is only correlational, and as such, we do not know the directionality of the effect. It seems plausible that some cultures have a specific view of beauty and success, which is then embedded in the corresponding language. However, it is equally possible that the language endogenously developed, leading to differential perceptions of beauty and success. To answer the causal question seems close to impossible at a large scale, but future research might be able to find causal empirical strategies identifying the direction of the effect. Sixth, our study focuses solely on providing evidence of a linguistic beauty premium. However, we do not address whether the relative strength of this premium predicts country-level outcomes, such as wages. Future research could explore this by identifying suitable variables recorded across a sufficiently large number of countries to extend our approach and examine whether and how the linguistic beauty premium translates into socio-economic outcomes.

Next to the suggested extensions, our study opens up other new avenues for future research. Our linguistic approach enables the simultaneous investigation of the existence and possible explanations of cultural influence on the beauty premium directly in real-world cultural traces at a relatively low cost. While we focused on detecting the linguistic beauty premium and existing explanations, our approach can serve as a blueprint for future research investigating new, culture-specific factors explaining the beauty premium. Training new word embeddings on corpora generated by a particular subgroup could also be used to study different linguistic beauty premia within the same language. Finally, our cross-lingual comparison provides hypotheses for a large number of regions without any prior information about the local extent of the beauty premium. Our data is, therefore, an invitation to test this hypothesis or to use it for further analyses.

Overall, we do not consider our study the last word on the matter but rather an initial piece of evidence on how culture might influence the beauty premium. This provides a direction for further research to create a more comprehensive understanding of the beauty premium.

Materials and methods

Data: pre-trained word embeddings

Our analysis utilizes pre-trained word embeddings (in particular, fasttext⁵²). Word embeddings are vector representations of words arranged such that the distance between any two vectors mirrors their semantic relationship. This arrangement considers the context of each word in the corpus, in line with the linguistics distribution hypothesis, implying words derive meaning from their context²⁹. The model iteratively adjusts vector distances based on word co-occurrence probability. These word embeddings are derived from extensive machine learning training on vast, diverse texts to reflect general linguistic features rather than specific textual properties³⁰. These models developed for more than 100 different languages³¹, can serve as inputs to models finetuned for specific tasks⁴⁷, and have demonstrated human-like notions of semantic relationships³⁰. The underlying pre-trained word embeddings we use were trained on a mixture of texts from Wikipedia and large sets of texts collected from the web in the respective language³¹. While the size and sources of the training data vary widely between languages (from around 1.9 billion words in Hindi to 650 billion words in English), it is unlikely to have a significant impact on our results as 1) even the smallest corpus is vast, 2) the methodology of text collection is consistent between languages, and 3) these embeddings have previously been shown to represent fundamental cultural differences, such as gender bias³⁷. Following the argumentation of previous research on bias in word embeddings, pre-trained models benefits our study by boosting replicability and representativeness due to adherence to best practices during training^{34,35,37}. Moreover, these models have been trained on a large amount of data, likely representing a wide variety of language use within a given culture.

Concept selection

To use word embeddings to study our objects of investigation – i.e., the beauty premium – we first identify key concepts of interest that represent the beauty premium based on the literature. The beauty premium broadly refers to the phenomenon that more physically attractive individuals receive certain benefits. We, therefore, replicate the beauty premium in word embeddings as the relationship between the concepts of beauty (represented by the antonym concepts *Beauty/Ugliness*) and success (*Success/Failure*) and affluence (*Affluence/Poverty*), respectively. Furthermore, the concepts of trust (*Trust/Distrust*), self-confidence (*Confident/Doubtful*), and competence (*Competence/Incompetence*) are selected to analyze the relationship between beauty and personal characteristics. Further, the positive concepts (*Good/Bad* and *Pleasant/Unpleasant*) have been used.

Word selection

As common practice, we use multiple related words to represent concepts^{34,35}. This approach captures various facets of a concept's meaning, reduces bias from single terms, and allows the construction of confidence intervals for word selection sensitivity. While word and concept selection relies on informed judgment, we enhance our method's robustness by using tested vocabulary from linguistic and psychological studies (see SI G for an overview of the vocabulary and sources). In order to study the beauty premium across languages, we first translated the selected words for each selected concept from English to the respective languages using an online translation tool. We then used these words in the respective languages to obtain the corresponding pre-trained word embeddings in these languages. We excluded regional accents and languages for which translation was not possible, resulting in word embeddings for 68 different languages.

POLAR framework

We build on the POLAR framework for capturing cultural associations in word embeddings^{40,53}. The POLAR framework is designed to add an interpretable dimension to the otherwise only relatively interpretable elements of pre-trained word embeddings. To illustrate, consider the linguistic beauty premium. We are interested in the association between the concepts *Beauty* and *Success*. We cannot obtain a meaningful interpretation of this association by a measure of similarity between these two concepts alone, as similarity can only be interpreted relative to other similarities. Therefore, we need to obtain the value of the concept *Beauty* on a *Success-Failure* dimension, i.e., is *Beauty* relatively more associated to *Success* or to *Failure* on a continuous *Success-Failure* scale. The POLAR framework offers such a score simultaneously for multiple vectors and dimensions⁴⁰.

As a first step, we select a set consisting of pairs of word vectors, each representing a pole on a dimension. For example, for the *Success-Failure* dimension, we consider { $(successful, unsuccessful), (victorious, failed), (winning, losing), ...}.$

By calculating the vector difference between the opposite word vector pairs, we obtain the direction vector for each antonym pair. These direction vectors can be stacked to form a direction matrix:

$$Dir^{Success/Failure} = \begin{cases} \overrightarrow{successful} - \overrightarrow{unsuccessful} \\ \overrightarrow{victorious} - \overrightarrow{failed} \\ \overrightarrow{winning} - \overrightarrow{losing} \\ \cdots \end{cases}$$

We can get the values of a set of words with respect to each of the direction vectors by projecting their corresponding vectors, e.g., { beautiful, attractive, handsome, ...}, onto all of the dimensions. The result is a single value

for each of the words of interest ($\{\overline{beautiful}, \overline{attractive}, \overline{handsome}, ...\}$) for all dimensions ($\{\overline{successful}, \overline{successful}, \overline{victorious}, -\overline{failed}, \overline{winning}, -\overline{losing}, ...\}$). A positive value of, e.g., $\overline{beautiful}$ projected onto the dimension $\overline{winning} - \overline{losing}$ can be interpreted as a stronger association of *beautiful* with *winning* than with *losing* and vice versa⁴⁰. Since, in our case, the different dimensions in the direction matrix and the vectors of interest all represent an overarching concept (i.e., *Success, Failure, Beauty*), we can simply aggregate these measures. By averaging the values over the different dimensions, we obtain a single score for each *Beauty*-word, representing its value on the *Success-Failure* dimension.

Finally, we average all these scores again to obtain a single score representing the value of the concept *Beauty* on the *Success-Failure* dimension, i.e., the linguistic beauty premium. By averaging over multiple word scores, we ensure that the vocabulary length for each concept does not bias the results. We obtain confidence intervals and the relevant statistics for *Beauty* on each of the dimensions by treating each *Beauty*-word as one independent observation. For each *Beauty*-word we can also obtain confidence intervals by mutating the word-antonym pair combination to construct the direction matrix. In order to obtain p-values, we use the average score of each word on the relevant dimension (e.g., the *Success-Failure* dimension) and compare, using t-tests, whether words representing *Beauty* differ, in this score, from words representing *Ugliness*.

The linguistic beauty premium on the country level

We use the POLAR framework to measure the linguistic beauty premium by placing *Beauty* on the *Success-Failure* dimension for each of the 68 languages we use. Thereafter, we associate each country with their primary spoken language (obtained from GeoNames.org). Using this approach gives us the linguistic beauty premium for most countries of the world.

Data availability

The data we use are pre-trained word embeddings freely available at https://fasttext.cc/.

Received: 12 August 2024; Accepted: 16 May 2025 Published online: 21 May 2025

References

- 1. Hamermesh, D. S. & Biddle, J. E. Beauty and the Labor Market. Am. Econ. Rev. 84, 1174-1194 (1994).
- 2. Hamermesh, D. S. & Parker, A. Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity. *Econ. Educ. Rev.* 24, 369–376 (2005).
- 3. Hamermesh, D. Changing looks and changing "discrimination": The beauty of economists. Econ. Lett. 93, 405-412 (2006).
- 4. Deryugina, T. & Shurchkov, O. Now you see it, now you don't: The vanishing beauty premium. J. Econ. Behav. Org. 116, 331–345 (2015).
- 5. Li, Y., Peng, L., Ma, S. & Zhou, X. Beauty premium or beauty penalty in sharing accommodation situations based on lay theories. *Int. J. Contemp. Hosp. Manag.* 34, 929–950 (2021).
- 6. Langlois, J. H. et al. Maxims or myths of beauty? A meta-analytic and theoretical review. Psychol. Bull. 126, 390-423 (2000).
- 7. Shapir, O. M. & Shtudiner, Z. Beauty is in the eye of the employer: Labor market discrimination of accountants. *Front. Psychol.* 13, 928451 (2022).
- Scholz, J. K. & Sicinski, K. Facial Attractiveness and Lifetime Earnings: Evidence from a Cohort Study. Rev. Econ. Stat. 97, 14–28 (2015).
- 9. Mobius, M. M. & Rosenblat, T. S. Why beauty matters. Am. Econ. Rev. 96, 222-235 (2006).
- 10. Dion, K., Berscheid, E. & Walster, E. What is beautiful is good. J. Pers. Soc. Psychol. 24, 285-290 (1972).
- 11. Jackson, L. A., Hunter, J. E. & Hodge, C. N. Physical attractiveness and intellectual competence: a meta-analytic review. Soc. Psychol. Q. 58, 108 (1995).
- 12. Ravina, E. An assessment of the mating motive explanation of the beauty premium in market-based settings. *Behav. Brain Sci.* 40, e39 (2017).
- Courtiol, A., Raymond, M., Godelle, B. & Ferdy, J.-B. Mate Choice and Human Stature: Homogamy as a Unified Framework for Understanding Mating Preferences. *Evolution* (2010).
- 14. Macchi, E. Worth your weight: Experimental evidence on the benefits of obesity in low-income countries. Am. Econ. Rev. (2023).
- 15. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world?. *Behav. Brain Sci.* 33, 61–83 (2010).
- Peng, L., Cui, G., Chung, Y. & Zheng, W. The faces of success: beauty and ugliness premiums in e-commerce platforms. J. Mark. 84, 67–85 (2020).
- 17. Deng, W., Li, D. & Zhou, D. Beauty and job accessibility: New evidence from a field experiment. J. Popul. Econ. 33, 1303–1341 (2020).
- 18. Henrich, J. P. The secret of our success (Princeton University Press, 2016).
- 19. Boyd, R. & Richerson, P. J. The Origin and Evolution of Cultures (Evolution and Cognition) (Oxford University Press, USA, 2004).
- 20. Boyd, R. & Richerson, P. J. Culture and the Evolutionary Process (University Of Chicago Press, 1988).
- 21. Gelfand, M. J. et al. Differences between tight and loose cultures: A 33-nation study. Science 332, 1100–1104 (2011).
- 22. Acemoglu, D. Obedience in the Labour Market and Social Mobility: A Socioeconomic Approach. Economica 89 (2022).
- 23. Acemoglu, D. & Jackson, M. O. Social Norms and the Enforcement of Laws. J. Eur. Econ. Assoc. 15, 245–295 (2017).
- 24. Barro, R. J. & McCleary, R. M. Religion and economic growth across countries. Am. Sociol. Rev. 68, 760 (2003).
- 25. Shiller, R. J. Narrative Economics. Am. Econ. Rev. 107, 967-1004 (2017).
- Dion, K. K., Pak, A.W.-P. & Dion, K. L. Stereotyping Physical Attractiveness: A Sociocultural Perspective. J. Cross Cult. Psychol. 21, 158–179 (1990).
- Shaffer, D. R., Crepaz, N. & Sun, C.-R. Physical Attractiveness Stereotyping in Cross-Cultural Perspective: Similarities and Differences between Americans and Taiwanese. J. Cross Cult. Psychol. 31, 557–582 (2000).
- 28. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient Estimation of Word Representations in Vector Space (2013).
- Jurafsky, D. & Martin, J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall Series in Artificial Intelligence (Pearson Prentice Hall, Upper Saddle River, N.J, 2023), 3nd ed edn.
- Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. Enriching word vectors with subword information. Trans. Assoc. Comput. Ling. 5, 135–146 (2017).
- 31. Grave, E., Bojanowski, P., Gupta, P., Joulin, A. & Mikolov, T. Learning word vectors for 157 languages (2018).

- 32. Gennaro, G. & Ash, E. Emotion and Reason in Political Language. Econ. J. 132, 1037-1059 (2022).
- 33. Ash, E., Chen, D. & Naidu, S. Ideas Have Consequences: The Impact of Law and Economics on American Justice. Q. J. Econ. forthcoming (2025).
- 34. Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183–186 (2017).
- Kozlowski, A. C., Taddy, M. & Evans, J. A. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. Am. Sociol. Rev. 84, 905–949 (2019).
- Garg, N., Schiebinger, L., Jurafsky, D. & Zou, J. Word embeddings quantify 100 years of gender and ethnic stereotypes. Proc. Natl. Acad. Sci. 115 (2018).
- 37. Lewis, M. & Lupyan, G. Gender stereotypes are reflected in the distributional structure of 25 languages. *Nat. Hum. Behav.* 4, 1021–1028 (2020).
- Greenwald, A. G., McGhee, D. E. & Schwartz, J. Measuring individual differences in implicit cognition: The implicit association test. J. Pers. Soc. Psychol. 74, 1464–1480 (1998).
- 39. Arseniev-Koehler, A. & Foster, J. G. Machine Learning as a Model for Cultural Learning: Teaching an Algorithm What it Means to be Fat. Sociol. Methods Res. 51, 1484–1539 (2022).
- Mathew, B., Sikdar, S., Lemmerich, F. & Strohmaier, M. The POLAR Framework: Polar Opposites Enable Interpretability of Pre-Trained Word Embeddings. In Proceedings of The Web Conference 2020, 1548–1558 (ACM, Taipei Taiwan, 2020).
- Eagly, A. H., Ashmore, R. D., Makhijani, M. G. & Longo, L. C. What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychol. Bull.* 110, 109–128 (1991).
- 42. Oosterhof, N. N. & Todorov, A. The functional basis of face evaluation. Proc. Natl. Acad. Sci. 105, 11087–11092 (2008).
- 43. Wilson, R. K. & Eckel, C. C. Judging a Book by its Cover: Beauty and Expectations in the Trust Game. *Polit. Res. Q.* 59, 189–202 (2006).
- 44. Todorov, A., Pakrashi, M. & Oosterhof, N. N. Evaluating Faces on Trustworthiness After Minimal Time Exposure. Soc. Cogn. 27, 813–833 (2009).
- Jones, A. L. & Kramer, R. S. Facial first impressions form two clusters representing approach-avoidance. Cogn. Psychol. 126, 101387 (2021).
- Jones, A. L., Shiramizu, V. & Jones, B. C. Decoding the language of first impressions: Comparing models of first impressions of faces derived from free-text descriptions and trait ratings. *British Journal of Psychology* (2024).
- Joulin, A., Grave, E., Bojanowski, P. & Mikolov, T. Bag of tricks for efficient text classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 427–431 (Association for Computational Linguistics, Valencia, Spain, 2017).
- Schnabel, T., Labutov, I., Mimno, D. & Joachims, T. Evaluation methods for unsupervised word embeddings. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 298–307 (Association for Computational Linguistics, Lisbon, Portugal, 2015).
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B. & Jetten, J. Out of mind but back in sight: Stereotypes on the rebound. J. Pers. Soc. Psychol. 67, 808–817 (1994).
- 50. Chen, M. K. The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets. Am. Econ. Rev. 103, 690-731 (2013).
- 51. Tabellini, G. Presidential Address Institutions and Culture. J. Eur. Econ. Assoc. 6, 255-294 (2008).
- 52. Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C. & Joulin, A. Advances in pre-training distributed word representations. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018) (2018).
- Engler, J., Sikdar, S. & Lutz, M. & Strohmaier, M (Word sense aware interpretability for pre-trained contextual word embeddings, Sensepolar, 2023).

Author contributions

B.K. and W.M. contributed equally to the paper.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/1 0.1038/s41598-025-02857-4.

Correspondence and requests for materials should be addressed to W.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025