Journal of Personality and Social Psychology

The Structure and History of English Words That Describe People

Yuanze Liu, Tessa Charlesworth, Alex Koch, Andrew Luttrell, and Joshua Conrad Jackson

Online First Publication, November 12, 2026. https://dx.doi.org/10.1037/pspp0000627

CITATION

Liu, Y., Charlesworth, T., Koch, A., Luttrell, A., & Jackson, J. C. (2026). The structure and history of English words that describe people. *Journal of Personality and Social Psychology*. Advance online publication. https://dx.doi.org/10.1037/pspp0000627

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Personality and Social Psychology:
Personality Processes and Individual Differences

© 2026 American Psychological Association
ISSN: 0022-3514

https://doi.org/10.1037/pspp0000627

The Structure and History of English Words That Describe People

Yuanze Liu[1], Tessa Charlesworth[2], Alex Koch[3], Andrew Luttrell[4], and Joshua Conrad Jackson[5, 6]

[1] Harvard Business School, Harvard University
[2] Kellogg School of Management, Northwestern University
[3] Booth School of Business, University of Chicago
[4] Department of Psychological Science, Ball State University
[5] Department of Psychology, Harvard University
[6] Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University

The English language is filled with words that are used to describe people, such as "nice," "skilled," "conservative," and "attractive." This broad trait space includes words that describe personality but also words that describe people's physical, sociodemographic, and ideological differences. Here we define trait space as the sum of stable, comparative, and socially meaningful descriptors and map this space by combining large language models, human ratings, and historical text analyses. Across four studies, we (a) introduce a comprehensive trait list (2,847 trait words); (b) establish normative human ratings of these words (3,070 English speakers) along 24 dimensions drawn from prior literature; (c) identify a four-factor structure of fitness, agency, communion, and traditionalism from these dimensions, which covers 95.4% of trait words; and (d) reveal systematic differences in the coherence, valence, and evolution of these factors. Specifically, these four factors differ in their coherence (e.g., communion is the most semantically coherent while traditionalism the least), valence (e.g., communion and agency show more alignment between their semantic direction and valence than traditionalism and fitness; communion has more negative words while agency and traditionalism more positive ones), and historical usage (e.g., fitness and communion are older than traditionalism; communion is converging in semantic space while agency diverging). These findings showcase new large language model–based methods for discovering and analyzing trait words. We close by motivating a functional constructionist view: The meaning, structure, and use of traits constitute a culturally constructed toolkit that represents socially relevant individual differences to help navigate recurring interaction problems.

Keywords: human traits, personality, social evaluation, lexical studies, large language models

Supplemental materials: https://doi.org/10.1037/pspp0000627.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

People love to talk about people. We talk about our family, friends, colleagues, strangers, and ourselves. People talk about people even when they are trying not to. The irony of the famous quote, "great minds discuss ideas; small minds discuss people," is that it is a comment about people. To facilitate talking about people, English contains thousands of words for describing human differences. These words include not only personality traits in the narrower dispositional sense (McAdams & Pals, 2006) but also other relatively stable, comparative, and socially meaningful characteristics, such as physical, sociodemographic, and ideological differences. For example, words such as warm, tall, wealthy, or conservative can all shape how a person is socially represented and evaluated. Together, these words form what we call trait[1]

[1] In contemporary personality psychology, the term *trait* is often used more narrowly to refer to relatively enduring differences in thought, feeling, and behavior (McAdams & Pals, 2006). We acknowledge that this is the dominant usage for many readers of personality psychology. At the same time, the history of the term is more heterogeneous. In their seminal lexical work, Allport and Odbert (1936) adopted the label *trait names* somewhat loosely across multiple columns of person-descriptive terms while explicitly noting that not all psychologists would confine the concept to its "strictest sense" (i.e., dispositional traits). More broadly, everyday language and other scholarly traditions, including biology and cultural anthropology, also often use trait in a wider sense (Dawson et al., 2021; Henrich & McElreath, 2003). In this article, we therefore use trait in this broader sense to refer to a wider space of relatively stable, socially meaningful ways of describing persons.

Kai T. Horstmann served as action editor.
Yuanze Liu https://orcid.org/0000-0003-1437-1892
Tessa Charlesworth https://orcid.org/0000-0001-5048-3088
Andrew Luttrell https://orcid.org/0000-0002-6210-4839
Yuanze Liu played a lead role in data curation, formal analysis, visualization, and writing–original draft and an equal role in conceptualization, methodology, and writing–review and editing. Tessa Charlesworth played a supporting role in conceptualization, methodology, and supervision and an equal role in writing–review and editing. Alex Koch played a supporting role in conceptualization, methodology, and supervision and an equal role in writing–review and editing. Andrew Luttrell played a supporting role in conceptualization, methodology, and supervision and an equal role in writing–review and editing. Joshua Conrad Jackson played a lead role in supervision, a supporting role in formal analysis and visualization, and an equal role in conceptualization, methodology, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Yuanze Liu, Harvard Business School, Harvard University, Soldiers Field, Boston, MA 02163, United States, or Joshua Conrad Jackson, Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, United States. Email: yuanzeliu@hbs.edu or jcjackson@fas.harvard.edu

space: a broader representational space through which people describe, evaluate, and make sense of others. Our goal in this article was to use recent innovations in natural language processing to systematically investigate this broader trait space by identifying relevant words and analyzing their semantic structure, valence, and historical evolution.

This article makes three contributions. The first two are methodological contributions. First, we develop a lexicon construction pipeline that combines large language model (LLM)-based classification with human validation, enabling efficient, transparent, and reproducible lexicon development at scale. Second, we use this pipeline to build a comprehensive lexicon of English human trait words, accompanied by extensive supporting information, including normative ratings on 24 trait dimensions from 3,070 native English speakers, as well as word-level metadata such as valence, frequency of use, and date of origination. We also provide an online interactive webpage (https://societyinmind.net/assets/trait-network/) for searching words and inspecting their metadata and network location. The third contribution is empirical and interpretive: Leveraging these two resources, we systematically characterize the (a) semantic structure, (b) evaluative properties, and (c) historical evolution of the trait space. We then use these patterns to motivate a functional constructionist framework, a framework that can guide new hypotheses for future research.

Human Trait Dimensions: Two Fields

Traits have become the center of two fields of psychology. The first, personality psychology, uses traits to describe enduring and stable individual differences in behaviors, thoughts, and feelings, often with the assumption that these individual differences are real, innate, and cross-culturally isomorphic (DeYoung et al., 2010; McAdams & Pals, 2006; McCrae & John, 1992). The second, social evaluation, uses traits to describe the stereotypes and impressions that people form of each other, with no assumption that these impressions are stable, real, or psychological in nature (Abele et al., 2021; Koch et al., 2021). Despite these differences, researchers in both fields try to identify low-dimensional models that explain variation in the myriad ways that humans can be described, and both fields have used lexical methods to develop their models.

Our goal in this article was to connect these traditions by studying the trait language that both fields rely on to describe individuals, using recent innovations in LLMs that enable trait space analysis at a new scale and with greater transparency. Because of our scope, we use the term "trait" in a slightly broader sense than is typical in personality research. Following McAdams and Pals's (2006) definition of traits as broad, nonconditional, decontextualized, generally linear and bipolar, and implicitly comparative dimensions of human individuality, we include any perceivable, relatively stable, and socially evaluative descriptors of individuals that can be located on bipolar continua, not only internal dispositions in behavior, thought, and feeling. Before introducing our work, we first briefly review the legacy of personality and social evaluation science and how these fields arrived at their seminal models.

The Psychology of Personality

Galton proposed that human character is a definite and durable "something" that can be measured and inherited (Galton, 1884, p. 181). He further proposed the famous lexical hypothesis: that it was possible to tap "the more conspicuous aspects of the character by counting in an appropriate dictionary the words used to express them" (Galton, 1884, p. 181). Galton estimated one thousand character words in the Thesaurus. Allport and Odbert (1936) then expanded this list by building a larger trait word list consisting of around 4,500 words from the Webster's Dictionary.

Thurstone (1934) extended research on personality by testing new questions about trait structure. He reported five factors of human traits from ratings of 60 trait adjectives. Cattell (1943) detected 12 factors based on factor analysis of 171 words, mostly from Allport and Odbert's (1936) list. Among Cattell's 12 factors were five factors that eventually became the "Big Five" model (Digman & Takemoto-Chock, 1981; McCrae & Costa, 1987; Norman, 1963): openness, conscientiousness, extraversion, neuroticism, and agreeableness (McCrae & Costa, 1987). The Big Five has achieved stunning popularity in personality science (Roberts & Yoon, 2022), and early supporters of the model have claimed that the Big Five is cross-culturally universal (McCrae et al., 1998) and may even characterize nonhuman animals (Gosling & John, 1999).

Subsequent lexical programs built on this tradition while proposing somewhat different structures. For example, the honesty–humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience model added a sixth honesty–humility factor (Ashton & Lee, 2007), and cross-cultural work on the "Big Two" summarized traits based on two dimensions of agency/dynamism and communion/social self-regulation (Thalmayer et al., 2025). Other work derived a "Big Seven" structure that retained the Big Five but added two broad evaluative dimensions (positive and negative valence; Benet-Martínez & Waller, 1997; Benet & Waller, 1995), hinting that trait language naturally blends dispositional content with social-evaluative meanings.

The Psychology of Social Evaluation

Asch (1946) famously pointed out that people possess a remarkable capacity to evaluate others' character quickly and automatically, with very little information. In turn, Asch wondered whether there were lawful principles that guided this evaluation (Asch, 1946). Asch's research focused on the role of the warm–cold dimension of impression formation (Asch, 1946). Rosenberg et al. (1968) later proposed and validated intellectual desirability as another basic dimension of social evaluation.

There are now several popular accounts of social evaluation applied in different contexts. These include the dual perspective model (Abele & Wojciszke, 2007), the behavioral regulation model (Ellemers et al., 2013), the dimensional compensation model (Yzerbyt et al., 2005), the stereotype content model (Fiske et al., 2002), and the agency, beliefs, and communion (ABC) model (Koch et al., 2016). The authors of these models recently synthesized their models into a single model with two broad dimensions (i.e., the Big Two): a vertical dimension that includes ability and assertiveness and a horizontal dimension that includes friendliness and morality (Abele et al., 2021; Koch et al., 2021; Yzerbyt et al., 2025). Versions of this Big Two model dominate the study of social evaluation in studies that rely on lexical measures (Abele et al., 2016; Cuddy et al., 2009), biopsychology (Harris & Fiske, 2006, 2007), and face perception (Oosterhof & Todorov, 2008). In this introduction, we refer to the Big Two as warmth/communion and competence/agency.

Envisioning a Broader View Into Trait Space

The fields of personality psychology and social evaluation both provide useful views into the structure of trait space. However, they

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

have focused on specific stimuli and a specific range of questions because of their theoretical goals. Here we explore what could be gained by taking a broader view of trait space and a functional perspective on why we develop words for this broader trait space.

The Scope of Trait Space

Research in personality psychology has made major progress in identifying the structure of personality traits, that is, relatively enduring differences in thought, feeling, and behavior (McAdams & Pals, 2006). Research on social evaluation has made parallel progress by identifying a small number of broad dimensions, such as agency and communion, that organize impression formation and social judgment (Abele et al., 2021). These traditions have generated valuable insights, but each does so by focusing on a restricted portion of trait space.

These boundaries may be sufficient for the traditional questions of each field. However, they are less complete for answering our core question: How do people guide social life by constructing a trait space composed of relatively stable, comparable, and socially meaningful human differences? By this standard, many other trait domains, such as physical, sociodemographic, and ideological differences, are also consequential for impression formation, social evaluation, and interaction (Eagly et al., 1991; Mallinas et al., 2018; Mattan et al., 2017). A more complete account of trait space therefore requires moving beyond both personality traits in the narrow dispositional sense and a small set of top-down evaluative dimensions toward the broader space of person description that people use in daily life.

The Function of Trait Space

Another question concerns what trait space reflects. Personality psychologists often treat trait space as if it were a direct and accurate representation of individual differences (see McCrae & Costa, 1997). However, this view overlooks how trait space contains systematic distortions based on what is useful to communicate. For example, in highly stratified cultures with tight social norms, like the Maasai of Eastern Africa, there are prominent dimensions reflecting age and power and no dimension of openness (Thalmayer et al., 2020). Cultural differences like these suggest that trait space is not a neutral map of all individual differences but a socially sculpted terrain that makes it easier for people to navigate their social world (Church, 2000; Triandis & Suh, 2002).

Social evaluation scholars accept this functional view of trait space as a tool for navigating the social world (Dupree & Fiske, 2017). However, classical models of social evaluation typically assume that there are dedicated and biologically determined mechanisms for how people perceive trait space, which downplay how trait space might actively change because of changes in our social ecology (Durkee et al., 2022; Montiglio et al., 2013). For example, the beliefs dimension of the ABC model (Koch et al., 2016) might have emerged alongside the rise of participatory democracies, as people's political ideologies became a more relevant part of social life (Jost et al., 2009).

A Functional Constructionist View of Trait Space: Three Questions

Our goal in this article was to gain a more complete view of trait space. Methodologically, we use fine-tuned large language models to identify a comprehensive list of English trait words in a more transparent, replicable, and scalable way. Conceptually, we adopt a functional constructionist perspective to generate research questions about the content, structure, and evolution of trait space. Like models of social evaluation, our perspective conceives of trait words as functional for person perception and prediction (Fiske, 1992). But like other constructionist approaches, we do not assume any deterministic structure of trait space (Russell & Barrett, 1999). Instead, we view trait space as actively constructed based on the structure of people's social ecology (Lindquist et al., 2022). We believe this functional–constructionist framework is broad enough to speak to dynamics of trait representations across all kinds of contexts, including cultures. However, in the present article, we empirically focus on diachronic change within the context of a single language (English).

Our perspective assumes that trait language has three key functions. First, trait words provide socially shared representations that bridge different contexts and behaviors, providing a structure to trait space. For example, a word like "friendly" provides a socially accessible schema for how someone will act in a range of different situations. Second, trait words are evaluative; words like "friendly" and "unfriendly" often communicate whether a person holds characteristics that are considered to be socially desirable or undesirable, helping humans position themselves and navigate partner-selection dilemmas. And third, trait words actively evolve to remain adaptive when social ecology changes. These three properties of trait language are aligned with what McAdams (1995) called an evolutionarily grounded scheme of "psychology of the stranger," which supports a quick, comparative readout of socially useful individual differences before we know details about their lives. The three properties also translate to three testable questions about trait words: (a) How is trait language semantically structured? (b) How does trait language encode positive or negative evaluation? (c) How does trait language evolve over time to reflect changing social ecology?

How Is Trait Language Semantically Structured?

The dimensional structure of trait language is probably the central question for the psychology of traits. The Big Five (and its six-factor variant of honesty–humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience) and Big Two are the most influential models of this structure, and they can be reconciled relatively easily (DeYoung, 2006; Digman, 1997; Peabody & Goldberg, 1989). For example, past studies have argued that agreeableness, conscientiousness, and neuroticism all load on a factor that roughly resembles warmth/communion, whereas extraversion and openness load on a factor that roughly resembles competence/agency. A similar two-factor model of personality shows relatively good generalizability across cultures (Thalmayer et al., 2025).

However, if we define traits more broadly as all perceivable and relatively stable individual differences that people use in social evaluations, then we may require a new model of trait space. Analyses of spontaneous evaluations have uncovered additional trait dimensions like gender, weight, and stylishness (Connor et al., 2026). Other articles argue that these are independent and meaningful dimensions that should be incorporated into models of person perception (Freeman & Lin, 2025; Hester & Hehman, 2023). As a whole, these classic and novel studies suggest that trait space may be organized into dominant dimensions resembling the Big Two or Big Five but may also include less dominant but nevertheless meaningful dimensions like gender and stylishness.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Beyond the number of dimensions, trait dimensions might vary in their semantic coherence. "Coherence" describes the strength of association among items in a category (Peabody & Goldberg, 1989). For example, if most words communicating morality also communicate friendliness, then this would suggest that friendliness and morality belong to a highly coherent overarching dimension. This structure guides people's inferences, leading people to predict that friendly peers should also be moral. Traditional studies of trait structure have focused more on the number of dimensions than the coherence of dimensions. However, from a functional perspective, coherence is also important for showing how people actively construct differentiated associations among trait words to make social predictions.

Some initial evidence suggests that warmth/communion dimensions may be more semantically coherent than competence/agency dimensions. For example, Eisenbruch and Krasnow (2022) have pointed out that people's warmth toward the same target is more consistent across time and context than their competence. This is because people's abilities are largely context specific (e.g., becoming athletic or well-read) and temporally variant (e.g., with maturing or aging), but different manifestations of warmth all reflect one general tendency—the extent to which people prioritize others' interest over their own (Delton & Robertson, 2012). Consistent with this argument, Jackson, Halberstadt, et al. (2023) found that people evaluated various moral traits, such as caring, respectful, and wholesome, as highly similar. However, this study did not compare morality to other dimensions, making variation in coherence across dimensions an open question.

How Is Trait Language Valenced?

If people use trait words to evaluate and compare people, then many of these words should be positively or negatively valenced. However, how should trait space be valenced to functionally serve our social life? We further unpack this broad question into two specific ones. The first question concerns the bipolar nature of trait dimensions. Every trait dimension has two poles, like the "friendly" and "unfriendly" poles of friendliness. In the case of friendliness, these poles are strongly and oppositely valenced (i.e., being friendly is positive, while being unfriendly is negative), but is this the case for all trait dimensions?

Many trait dimensions may correlate with valence because their poles align with basic human needs. For example, warmth/communion corresponds to the need for intimacy/acceptance/relatedness, whereas competence/agency corresponds to the need for power/competence (Dweck, 2017; McAdams, 1988; Ryan & Deci, 2000). Because people generally desire the satisfaction of basic needs and dislike the dissatisfaction of them, people may distribute valence to the two poles of trait dimensions that align with these basic needs. However, other trait dimensions may have more culturally or personally contingent value structures. For example, people likely evaluate beliefs (e.g., progressive vs. conservative) and extraversion (e.g., reserved vs. extraverted) differently based on their personal tastes and cultural background instead of assigning uniform valence to their semantic poles. No studies have exhaustively tested how the alignment between semantic direction and valence varies across trait dimensions.

The second question concerns the number of positive and negative trait words in each dimension—regardless of whether this dimension has oppositely valenced poles. Is our trait lexicon dominated by negative words? By positive words? Or does the relative number of negative and positive words vary across dimensions?

We see two competing hypotheses concerning the distribution of positive and negative trait words. The first hypothesis is that all trait dimensions have more negative than positive words. This hypothesis draws on research showing that negative information is more salient and consequential than positive information (Averill, 1980; Baumeister et al., 2001). People may therefore be more motivated to develop new words expressing and delineating negative information compared with positive information. A separate line of research further suggests that negative information is more heterogeneous than positive information (Alves et al., 2017). While positive traits often reflect moderate or desirable states, negative traits span a wider range of extremes. For example, extreme levels of a typically positive trait such as agreeableness may represent characteristics that are undesirable in different ways, such as showing excessive compliance (conformist) or open antagonism (quarrelsome), both of which deviate from the desirable level of agreeableness.

A rival hypothesis is that the balance of positive and negative trait words varies by dimension. For example, there may be more negative words about warmth/communion and more positive words about competence/agency. This hypothesis draws on research suggesting that negative traits are more diagnostic for morality, whereas positive traits are more diagnostic for competence (Skowronski & Carlston, 1987, 1989). Two complementary theories explain this asymmetry. The first is base-rate expectations (Kawamura & Barclay, 2025; Unkelbach et al., 2019, 2021): Most people are generally honest, so dishonesty is more informative, whereas few people are highly competent, making competence more informative. The second is expected net utility. Error management theory (Haselton & Buss, 2000; Johnson et al., 2013) suggests that the cost of misjudging others is asymmetric across domains: When it comes to honesty, mistaking a dishonest person for an honest one typically carries higher costs than mistakenly treating an honest person as dishonest; when it comes to competence, overlooking someone who is truly competent can be more consequential than overestimating someone who is incapable. Together, these factors predict a greater lexical emphasis on negative traits for warmth/communion and on positive words for competence/agency.

How Has Trait Language Changed Over Time?

Our third question concerns whether the structure of trait space has changed over time. The English language traces back to the 5th and 6th centuries Common Era and has since undergone continual evolution and expansion (Jackson, Lindquist, et al., 2023). Traditional universalist theories of traits would suggest that the structure of trait space has remained stable over this time. According to these views, traits reflect innate, biologically hardwired aspects of human nature—dimensions that humans are evolutionarily prepared to perceive, express, and evaluate in consistent ways across history (Dupree & Fiske, 2017; McCrae & Costa, 1997). We would therefore expect little historical variation in the predominance or semantic structure of trait dimensions. By contrast, a functional constructionist perspective suggests that historical changes in social ecology should lead people to actively reconstruct trait space, which should be reflected in significant changes in trait language.

There are two threads of research that make concrete hypotheses about how the semantic space of traits may evolve. The first thread

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

of studies has noted that the dimensional structure of traits varies across cultures (Aghababaei et al., 2025; Cheung et al., 2001; Gurven et al., 2013; Thalmayer et al., 2020, 2021). An emergentist perspective interprets this variation as evidence that the structure of personality emerges from people's social ecology and that the number and nature of trait dimensions should correspond to the number and nature of socioecological niches (Durkee et al., 2022; Montiglio et al., 2013; Smaldino et al., 2019). For example, the emergence of occupations (e.g., professor) that require new combinations of skills (e.g., public speaking, breadth of knowledge) may have given rise to new words that communicate these new skills (e.g., eloquent), expanding existing dimensions or creating new dimensions of competence.

A second line of research proposes that the evolution of trait space can also be shaped by the changing information ecology and constraints on human cognition. For example, Jackson, Halberstadt, et al. (2023) argued that as social networks become larger, more abstract trait representations may become functional heuristics for making social predictions with little information about interaction partners. We may now rely more on inference heuristics, rather than observations, in making judgments about others' traits (Crockett et al., 2024). In support of this argument, Jackson, Halberstadt, et al. (2023) found that morality trait space became more semantically coherent from 1800 to 2000. However, there is an open empirical question about whether all dimensions have become more coherent over history or whether this trend is unique to morality.

The Current Research: Four Studies

We explore our three core questions in four studies. Study 1 is devoted to developing our trait list, and Studies 2–4 are devoted to analyzing it.

In Study 1, we carefully curate one of the most comprehensive lists of English-language trait words to date. We start with the full English-language vocabulary of the Google Books from 1800 to 2008. We next apply a multistep process in which we first fine-tune three LLMs to recognize possible trait words from our full starting vocabulary and then use another LLM (ChatGPT) to remove words from this revised list whose primary meanings are not traits. Our resulting list includes 2,847 trait words that we analyze in the following studies. To facilitate transparency and exploration of the trait lexicon, we also provide an online interactive webpage wherein readers can search for individual words and inspect their associated metadata and network location (https://societyinmind.net/assets/trait-network/).

Study 2 analyzes the semantic structure of trait words. We begin by norming our complete 2,847-word trait list using a sample of 3,070 English speakers. Each participant rates trait words on a preregistered set of 24 dimensions from personality and social evaluation literature. We test whether these 24 dimensions are sufficient to capture the trait words in our full list, and we explore the semantic structure of trait words by analyzing participant ratings of each dimension using exploratory factor analysis (EFA). This EFA returns a four-factor model comprising fitness, agency, communion, and traditionalism (FACT). Further analyses test the coherence of the four FACT factors, finding that communion is the most coherent, while traditionalism is the least.

Study 3 next analyzes the valence of trait words. We test how differently valenced words are distributed within each dimension and which trait dimensions have more negative or positive trait words. Our analyses show that the distribution of valence correlates highly with the semantic direction of communion, agency, and fitness (except for the weight dimension) but does not map consistently onto the semantic direction of traditionalism. We also found that negative words are more fine-grained than positive words—communicating a narrower set of dimensions. However, this specificity does not uniformly translate to more negative words than positive words. Instead, there are more negative words in the communion dimensions and more positive words in the traditionalism and agency dimensions.

Finally, Study 4 analyzes the evolution of trait words. Using data from the Oxford English Dictionary, we estimate approximate dates of origin for trait words in our sample. We show that the relative frequency of trait words across dimensions has been remarkably stable over time. For example, there have always been more words about communion and agency than other factors, and communion and agency words have always been more frequently used than words in the other factors. However, we also found evidence of changing semantics and frequency over the last 200 years. For example, traditionalism words have become more commonly used over time, whereas communion words have become less frequent. We also show historical changes in trait semantic structure, indicating that communion has become more semantically coherent over the last 200 years relative to the other FACT factors.

Transparency and Openness

Analyses reported in Studies 2–4 were preregistered (see https://aspredicted.org/mftt-ngjq.pdf). Any deviations from our preregistration are explained in Supplemental Material 1. Data were analyzed using Python (model fine-tuning) and R (other data analyses). The full trait word list, normative rating data, and replication code to support transparency and reproducibility are available at https://osf.io/7k4yj/?view_only=8355ec82071e4b7e9d3cf6c11e1306e5.

Ethics Statement

This research was approved by the Social and Behavioral Sciences Institutional Review Board at the University of Chicago (Protocol IRB23-1428). The study was reviewed under the 2018 Common Rule and determined to be minimal risk. Approval was granted on October 20, 2023. All participants provided informed consent through an approved written consent form or information sheet in accordance with the Institutional Review Board policy.

Study 1

The purpose of Study 1 was to build a representative and extensive list of trait words. To be clear, we recognize that this trait list is not, and could never be, totally exhaustive. One reason is that trait words are natural categories (Rosch, 1973), meaning that there is no clear delineation between trait words and nontrait words. Moreover, the composition of trait language changes across time due to semantic drift and the invention and extinction of words (Hamilton et al., 2018). Hence, the comprehensiveness of trait lists is time-sensitive. As a case in point, Allport and Odbert's list contained many words that could plausibly be trait words but are mostly used in trait-unrelated domains or words that were once used to describe

human traits but are no longer, such as "blazer," "buck," and "Beau Brummell." We take a more conservative approach by focusing on words that convey traits as their primary meaning and are stably used from 1800 to 2008. To reduce the influence of researcher degrees of freedom and to facilitate efficient classification of the full English vocabulary, we used LLMs (Abdurahman, Atari, et al., 2024) with validation from human-subject ratings. Our study followed a two-step process. First, we fine-tuned three open-source pretrained bidirectional encoder representations from transformers (BERT) models (Devlin et al., 2019) on the English vocabulary of the Google Books corpus (Yin, 2018) to classify whether a word is or is not a trait word. After getting this preliminary list, we further refined it using a commercial, but more advanced, LLM: ChatGPT-4o (OpenAI, 2024).

Method

Materials

Source Vocabulary. To build an extensive list of trait words, we began with the full vocabulary of the Google Books corpus between 1800 and 2008. We started from a preprocessed list with words that were used no less than 100 times (Yin, 2018), which included 83,662 words. We intentionally chose this vocabulary because English vocabulary is constantly evolving, making its composition time-sensitive. Selecting words that have been stably used over a longer period (e.g., 1800s to 2008) yields more reliable results. However, this choice may exclude some recently popular but conceptually relevant words. We therefore encourage readers to keep this temporal sensitivity in mind when interpreting our results. We further preprocessed it by removing English stop words (e.g., functional words like "a," "the," "or," "would," and "you"), number words, and misspellings. This left us with a list of 53,284 words that we then used for LLM classification.

Training Set. Fine-tuning LLMs involves teaching the models about a more specific set of rules than they learned in their initial training. For our purpose, we wanted to teach LLMs how to tell the difference between trait words and nontrait words. To build a classification model like this, we needed to create a "training set" with examples of trait words and nontrait words, labeled as such.

Our examples of trait words in the training data set came from the classic Peabody list of representative trait words (Peabody, 1987), which is composed of 627 trait words. We decided to manually relabel some of these words as "nontrait" because they were (a) polysemous words whose primary meanings do not communicate traits (e.g., soft, direct) and (b) hyphenated compounds that communicate traits, but could not be processed in several of our analyses, and which we therefore wanted to exclude from our data set (e.g., self-critical, self-assured). Incorporating the Peabody list gave us 581 trait words and 46 nontrait words. To fill in our sample of nontrait words and get a more balanced training set, we randomly selected 688 words from our Google Books vocabulary. We manually labeled 44 of these as trait words and 644 as nontrait words. In total, this resulted in a training data set of 1,315 words: 625 trait words (581 from the Peabody list and 44 additional words) and 690 nontrait words (46 from the Peabody list and 644 additional words).

We appended our word list with Wiktionary definitions whenever they were available as learning material, hoping to provide models with richer information that they could use for classification. In the end, our training set consisted of a list of 1,315 words with their definitions and a column noting whether they are a trait word or not.

LLMs. We chose three BERT models for fine-tuning. BERT is a state-of-the-art pretrained language model specifically designed to capture the bidirectional context of a word. It produces more dynamic, context-dependent embeddings, which translates to more accurate word and text representations (Devlin et al., 2019). This makes BERT especially powerful for classification tasks, as its deep, contextual understanding allows it to excel at detecting minimal semantic changes. For example, while word2vec would assign the same representation to the word "warm" regardless of whether it appears in the context of "warm weather" or "warm personality," BERT distinguishes between these meanings by considering the surrounding words.

There are different versions of BERT models pretrained on different corpora with different parameters. We chose three of them: bert-base-uncased, bert-large-uncased, and Beijing Academy of Artificial Intelligence-bge-large-en-v1.5, which range from 110 million to 335 million parameters. We chose these three models because (a) they are open source, which facilitates replication; (b) previous research suggests that BERT models of this size can already outperform some commercial LLMs like ChatGPT 3.5 after fine-tuning (Abdurahman, Atari, et al., 2024); and (c) they are also widely used in the study of social psychological concepts (Bao, 2024; Cutler & Condon, 2023).

Procedure and Results

Learning Phase

LLM fine-tuning resembles human learning, involving phases such as learning, feedback, and testing. In our research, the learning and feedback phases are based on the annotated training set mentioned earlier. We split the training set into an 80:20 ratio, with 80% used for learning (learning set) and 20% for evaluation (feedback set). BERT's learning is based on a loss function, which measures the difference between the model's predictions and the actual labels and adjusts the model parameters accordingly to minimize this difference (Kozlowski et al., 2024). This learning process is typically repeated multiple times, referred to as different epochs. In each epoch, performance is measured using the F1 score, which is the harmonic mean of precision (the proportion of predicted positives that are true positives) and recall (the proportion of actual positives that are correctly identified), effectively balancing both. We saved the fine-tuned version of the model after each epoch and selected the version that performed best across 10 epochs for the subsequent formal classification tasks.

The F1 scores for the best performing versions of the three BERT models were satisfactory: F1 score = 0.94 for bert-base-uncased in Epoch 8, F1 score = 0.94 for bert-large-uncased in Epoch 7, and F1 score = 0.93 for Beijing Academy of Artificial Intelligence-bge-large-en-v1.5 in Epoch 3. These scores indicate excellent performance, as an F1 score close to 1.0 reflects a near-perfect balance between precision and recall, which is desirable in classification tasks. These are displayed in Panel A of Figure 1.

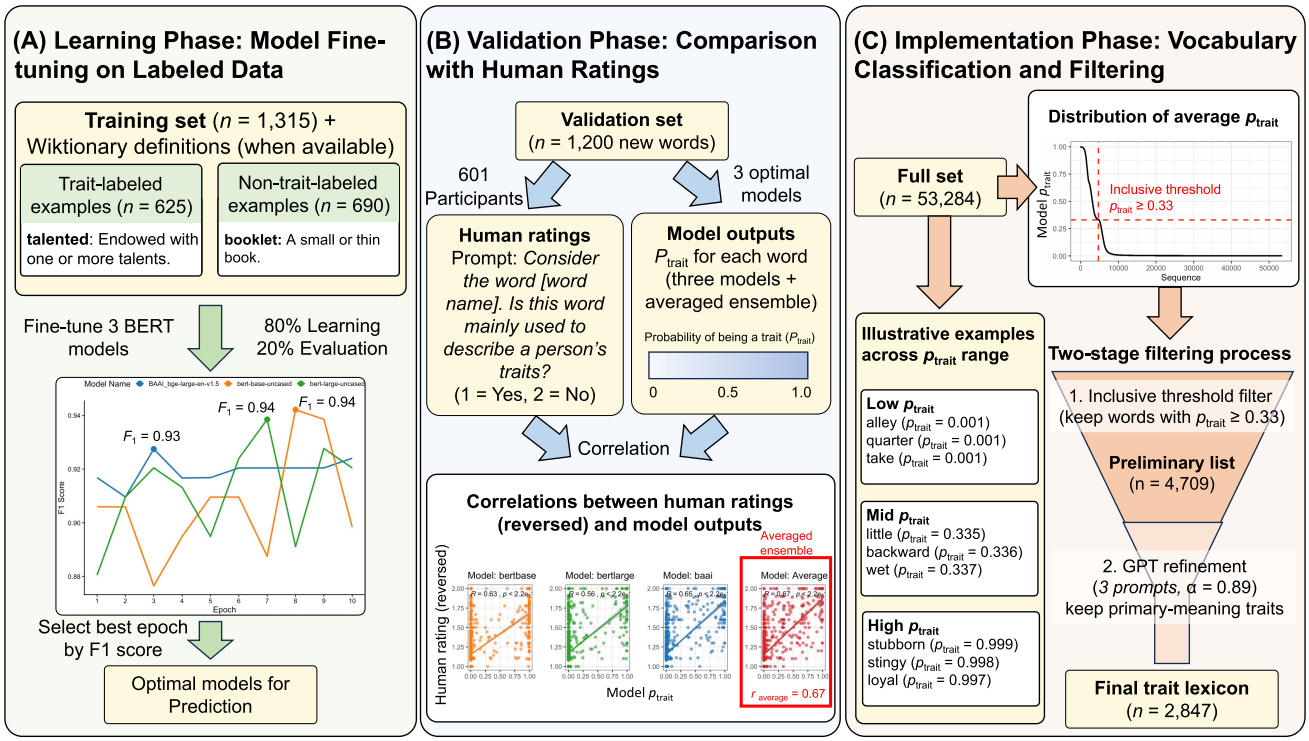
Validation Phase

After we fine-tuned the models, we further validated their effectiveness by comparing the best performing fine-tuned models with human ratings on a new sample of words. To do so, we randomly

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 1

Simplified Flowchart of the Trait Word List Development in Study 1



Note. (A) Learning phase: The training set (n = 1,315) is split into a learning set (80%) and an evaluation set (20%) to fine-tune three pretrained BERT models: BAAI-large, bert-base-uncased, and bert-large-uncased. The performance of the models is tracked across 10 epochs using the F1 score, shown in the bottom-left line graph. For each of the three BERT models, we selected the epoch-specific version with the highest F1 score (optimal models) in subsequent analyses. (B) Validation phase: The validation set (n = 1,200) is used to assess the performance of the models against manual ratings provided by human evaluators (n = 601). Scatterplots display correlations (r) between model ratings and manual ratings for each individual model (e.g., bert-base, bert-large, BAAI-large) and their averaged predictions, with high statistical significance ($p < .001$). (C) Implementation phase: The full data set (n = 53,284) is processed using the refined BERT ensemble model to create a preliminary trait word list (n = 4,709) based on a threshold probability (red dashed line), which is then filtered using ChatGPT judgments of whether each word's primary meaning describes a trait to produce the final refined list (n = 2,847). This filtering process was designed to retain words whose primary meanings describe traits. BERT = bidirectional encoder representations from transformers; BAAI = Beijing Academy of Artificial Intelligence. See the online article for the color version of this figure.

sampled 1,200 new words from the full preprocessed vocabulary. We had all of these words (and their Wiktionary definitions) evaluated by the three BERT models that performed best during the training phase, and we saved the binary classifications and underlying probabilities from each of these models. We recruited 601 Prolific workers (M_{age} = 39.53, SD = 12.10, 253 men, 342 women, and six chose the option "other genders"). The five most represented nations in our sample were the United Kingdom (n = 410, 68.22%), Canada (n = 70, 11.65%), the United States (n = 30, 4.99%), Australia (n = 20, 3.33%), and New Zealand (n = 15, 2.50%). These five countries comprised 90.68% of our participants. In terms of education, the largest share of participants (n = 247, 41.10%) earned an undergraduate degree, followed by those who earned a high school diploma (n = 169, 28.12%), a master's degree (n = 90, 14.98%), and an associate degree (n = 45, 7.49%). We asked participants to rate each word using the prompt: "Consider the word [word name]. Is this word mainly used to describe a person's traits?" (1 = yes, 2 = no, 3 = I don't know this word). Each person was randomly given 20 words to rate, and each word was rated by eight to 13 people. We used the average score of all raters as the manual rating of each word, turning ratings of "3" into missing values. We then compared the model ratings and manual ratings by calculating their correlation.

All three models showed good correlations with our human raters in their probability scores of being a trait for each word in the validation set: r = .63, $p < .001$ for bert-base-uncased; r = .56, $p < .001$ for bert-large-uncased; and r = .65, $p < .001$ for Beijing Academy of Artificial Intelligence-bge-large-en-v1.5. When averaged across models, the average probability scores also correlated significantly with human ratings, r = .67, $p < .001$. This correlation is comparable to that between LLMs' ratings and human ratings in other psychological constructs (i.e., .59–.77) in previous studies (Rathje et al., 2024). Readers should also note that even when individual raters show moderate-to-high agreement (e.g., interrater r = .80–.90), basic measurement theory implies that any new measure of the same construct will typically correlate in the .60–.80 range with that aggregate at best (Muchinsky, 1996). In our case, the gap is further amplified by the fact that the fine-tuned LLMs were trained on a binary classification objective (trait vs. nontrait), whereas human ratings reflect a more graded, continuous sense of "traitlikeness," which naturally produces some systematic discrepancies even when the underlying concept is shared.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Implementation Phase

After validating the effectiveness of model ratings, we classified the full set of 53,284 words in our vocabulary. In this way, we got the continuous probabilities of being a trait for each word. The main challenge at this point was to navigate the trade-off between insufficient recall (excluding true trait words from the list) and insufficient precision (retaining nontrait words in the list).

We took a two-step approach to addressing this challenge. In our first step, we prioritized inclusion, retaining all words that had a 0.33 probability or higher according to our classifiers. We selected a threshold of 0.33 because plotting the distribution of probabilities showed an elbow at 0.33, suggesting that there might be a meaningful difference between trait words with greater versus lesser than a 0.33 probability according to our BERT models. The threshold of 0.33 is also meaningful because it would be the probability of a word rated as a trait by one model, but not the other two. Applying this threshold produced a preliminary list of 4,709 words; nontrait words such as "alley" (p_{trait} = .001), "quarter" (p_{trait} = .001), and "take" (p_{trait} = .001) were removed.

We reasoned that this list probably contained a high volume of false positives. To find and exclude false positives, we used a commercial LLM: ChatGPT. ChatGPT is based on the generative pretrained transformer (GPT) model. Like BERT models, ChatGPT also generates dynamic, context-aware representations, allowing it to better capture the meaning of words in varying contexts and perform more complex language tasks (Radford et al., 2018). We used ChatGPT for this validation step for two main reasons. First, combining ratings from different LLMs can yield more accurate classifications than relying on one kind of LLM alone (Barabucci et al., 2024; Gu et al., 2024; Jiang et al., 2023). Second, GPT has highly flexible prompts, which makes it good for complex tasks like inferring the primary versus secondary meaning of words (Rathje et al., 2024). We used the "gpt-4o" application programming interface and set the temperature to zero, which means the output to the same query would have the lowest possible variance. We used three versions of the same prompt differing slightly in wordings to ask ChatGPT to assess if the primary meaning of a given word describes a person's trait (1 = yes, 0 = no). The responses to our prompts correlated highly (Cronbach's α = .89).

ChatGPT ratings averaged from three separate queries showed a high correlation with BERT-rated probabilities, r = .50, p < .001. Interestingly, this correlation was slightly lower than the corresponding correlations between BERT and human raters in the validation phase. We think this is partly because the response style of ChatGPT is less variable among prompts (Abdurahman, Ziabari, et al., 2024), thus constraining the correlation, whereas variation across human ratings could reflect a "wisdom of crowds." We retained words that were rated as trait words at least once across the three GPT queries to obtain our final list of trait words. This resulted in a final list of 2,847 trait words. Words removed included "little" (p_{trait} = .335), "backward" (p_{trait} = .336), and "wet" (p_{trait} = .337). Figure 1 illustrates the full procedure. Figure 2 illustrates these words, with word size representing the number of WordNet senses, which were extracted from WordNet (M = 3.28, SD = 3.49).

Discussion

Study 1 confirmed the feasibility of detecting trait words from a large English vocabulary using fine-tuned LLMs. Although no classification method is perfect, BERT models allowed us to classify a very large sample of 53,284 words using clear and consistent definitions of human traits and classification criteria. The whole classification task was performed based on criteria the models "learned" from the predefined training set during the fine-tuning phase and then stored as model parameters. This process makes our training process transparent and reproducible.

We acknowledge that our operational definition of trait words based on the Peabody list is not completely bias-free, but no list could be. Compared with alternative operationalizations, the Peabody list is designed to be representative (Peabody, 1987), and it provides a clear and explicit starting point for trait classification.

Another strength of our approach is that we obtained not only binary classifications but also continuous probability scores ranging from 0 to 1, which better reflect the nature of human traits as natural concepts. By "natural concepts," we mean that trait words have no discrete, clear-cut boundaries but rather fuzzy peripheries consisting of less prototypical cases (Rosch, 1973; Sainsbury, 1996). Although we speak about trait words as if they represented a fixed population, many words can be used to describe traits in some contexts but can be used in other ways in other contexts. These nonprototypical cases are polysemous words at the periphery of the trait word cloud in Figure 2. Because our approach provides a continuous probability score for each word in a large vocabulary (n = 53,284), future studies can adopt their own thresholds for trait categorization to reduce either false positives or false negatives.

Study 2

Study 2 examined the dimensional structure of the English trait lexicon based on the representative and extensive trait word list that we developed in Study 1. Specifically, we sought to answer three broad questions concerning dimensional structure: (a) How many trait dimensions are needed to cover the English trait lexicon? (b) Is there a parsimonious structure of these dimensions? (c) How coherently are these dimensions organized in the trait space? To these ends, we used an approach that combined top-down and bottom-up elements. We began with a top-down literature review of nonredundant trait dimensions, which we drew from the fields of personality psychology and social evaluation. Next, we recruited human raters to evaluate if each word from our trait list communicates each candidate dimension. Using information about which trait dimensions were commonly coexpressed by words, we could identify trait structure through exploratory dimension reduction techniques. Furthermore, we could analyze the semantic coherence of the resulting factors, testing whether some factors comprise looser dimensional associations than others.

Method

Materials

Trait Word List. We used the 2,847-word list that we developed in Study 1. We also use this list as the trait word list in all subsequent studies.

Candidate Dimensions. We reviewed three major literatures that measure human traits, summarized in Table 1. The first literature is personality research, including studies of the Big Five (Goldberg, 1990; McCrae & Costa, 1987) and the honesty–humility, emotionality, extraversion, agreeableness, conscientiousness, and

Figure 2

Word Cloud of Our Final Word List

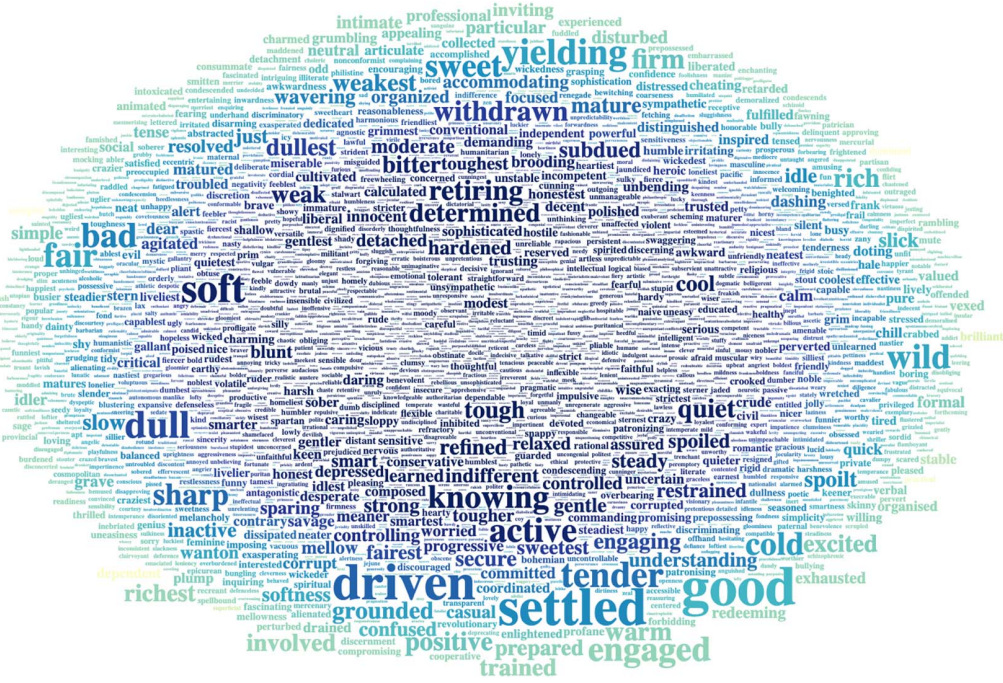

yielding
withdrawn
driven
settled
good
fair
soft
dull
slow
bad
sharp
knowing
active
retiring
determined
hardened
detached
weak
bitter
dullest
resolved
sweet
firm
mature
wild
cool
quiet
tough
refined
indifferent
controlled
tender
engaged
warm
trained
positive
involved
grounded
richest
inactive
wavering
weakest
accommodating
subdued
rich
idle
dashing
frank
slick
pure
calm
chill
formal
spoilt
cold
excited
understanding
secure
fairest
mellow
corrupt
softness
confused
drained
perturbed
exhausted
redeeming
prepared
enlightened
profane
forbidding
coordinated
committed
unbending
honest
brooding
heartiest
moderate
demanding
conventional
grimmest
just
tense
abstracted
resolved
matured
troubled
alert
agitated
dear
gentlest
shaky
shallow
brave
innocent
liberal
calculated
cultivated
unstable
incompetent
cunning
decent
polished
trusted
hostile
sophisticated
trusting
reserved
swaggering
awkward
intelligent
violent
naive
uneasy
educated
healthy
serious
hardly
religious
stout
coolest
effective
valued
vexed
brilliant
crabbed
perverted
unlearned
stressed
grim
incapable
versed
busy
silent
tenderness
doting
neatest
unfriendly
hale
happier
mute
zany
frail
virtuous
offended
lively
capable
dashing
idle
rich
fulfilled
fawning
retarded
cheating
distressed
sympathetic
receptive
demoralized
fulfilled
inspired
tensed
irritating
prosperous
frightened
mercurial
patrician
delinquent
approving
bully
honorable
distinguished
humble
heroic
lonelier
wickedest
masculine
arrogant
sagest
partisan
amusing
welcoming
benighted
outraged
flirt
chastened
insecure
pacific
informed
kindest
fierce
sulky
jaundiced
unfeeling
unmanageable
fetching
haughtiness
mercenary
patrician
honorable
bully
demoralized
condescends
fulfilled
fawning
retarded
cheating
distressed
sympathetic
receptive
mature
withdrawn
yielding
firm
sweet
particular
professional
inviting
intimate
appealing
collected
disturbed
neutral
articulate
accomplished
liberated
embarrassed
enchanting
confidence
cheating
distressed
sophistication
grasping
wickedness
focused
renegade
bewitching
coarseness
humiliated
weakest
accommodating
organized
wavering
indifference
unpredictability
haughtiness
fetching
sweetheart
reasonableness
harmonious
friendliest
dedicated
disarming
exasperated
irritated
lettered
sorrowing
abstracted
just
tense
grimmest
conventional
agnostic
independent
powerful
sensitiveness
inspired
moderate
demanding
humanitarian
lonely
strident
lawful
bitter
toughest
brooding
heartiest
moral
wickedest
masculine
arrogant
subdued
irritating
prosperous
forboring
frightened
mercurial
patrician
delinquent
approving
bully
honorable
distinguished
resolved
dullest
miserable
misguided
feebler
cordial
cultivated
freewheeling
concerned
unstable
incompetent
cunning
unbending
heroic
lonelier
wickedest
masculine
arrogant
sagest
partisan
amusing
welcoming
benighted
outraged
flirt
chastened
matured
deliberate
negativity
discretion
weak
stalwart
calculated
retiring
honest
outgoing
unfeeling
unmanageable
fetching
haughtiness
mercenary
patrician
honorable
bully
demoralized
condescends
troubled
feebler
showy
immature
stricter
liberal
innocent
determined
decent
polished
trusted
petty
nicest
silent
busy
frail
virtuous
offended
lively
capable
dashing
versed
frank
alert
brave
shallow
gentlest
shaky
detached
hardened
reserved
swaggering
awkward
intelligent
violent
naive
uneasy
educated
healthy
serious
hardly
religious
stout
coolest
effective
valued
vexed
brilliant
crabbed
perverted
unlearned
stressed
grim
incapable
versed
busy
silent
tenderness
doting
neatest
unfriendly
hale
happier
mute
zany
frail
virtuous
offended
lively
capable
dashing
idle
rich

Table 1
Candidate Trait Dimensions Used in the Present Study

No.	Dimension	Source	Anchor	Dimension removed
1.	Assertiveness	The Big Two (Abele et al., 2021)	Insecure/indecisive, confident/determined	
2.	Ability		Incompetent/unskilled, skilled/capable	
3.	Friendliness		Unfriendly/cold, warm/friendly	
4.	Morality		Immoral/unethical, moral/ethical	
5.	Conservatism	The ABC (Koch et al., 2016)	Progressive/liberal, conservative/traditional	Agency and communion are duplicated in the Big Two dimensions, thus were not included.
6.	Agreeableness	The Big Five (McCrae & John, 1992)	Critical/quarrelsome, sympathetic/agreeable	
7.	Conscientiousness		Disorganized/careless, dependable/self-disciplined	
8.	Openness		Conventional/uncreative, open to new experiences/complex	
9.	Extraversion		Reserved/quiet, extraverted/enthusiastic	
10.	Neuroticism		Calm/emotionally stable, anxious/easily upset	
11.	Honesty	HEXACO (Ashton & Lee, 2007)	Dishonest/insincere, honest/sincere	Other HEXACO dimensions strongly overlap with dimensions in the Big Five.
12.	Gender	Based on spontaneous descriptions of Facebook profile photos (Connor et al., 2026)	Feminine/female, masculine/male	Four dimensions are duplicated in the four facets of the Big Two model; conservatism dimension is duplicated with the ABC model; race dimension is multipolar.
13.	Age		Old/senior, young/junior	
14.	Adventurousness		Risk averse/indoorsy, adventurous/outdoorsy	
15.	Stylishness		Unstylish/unfashionable, stylish/fashionable	
16.	Fitness		Unfit/unathletic, athletic/fit	
17.	Nonconformity		Normal/ordinary, weird/unique	
18.	Attractiveness		Ugly/unattractive, attractive/beautiful	
19.	Weight		Underweight/thin, overweight/heavy	
20.	Status	Based on spontaneous labels of social groups (Nicolas et al., 2022)	Disreputable/low status, prestigious/high status	Four dimensions are duplicated in the four facets of the Big Two model; appearance is duplicated in attractiveness above; deviance is duplicated in nonconformity above; emotion, occupation, social group, and geography dimensions are multipolar.
21.	Power		Powerless/submissive, powerful/dominant	
22.	Wealth		Poor/destitute, rich/wealthy	
23.	Health		Unhealthy/sick, healthy/vigorous	
24.	Belief		Nonreligious/unspiritual, religious/spiritual	

Note. ABC = agency, beliefs, and communion; HEXACO = honesty–humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience.

Procedure and Results

Each participant rated 10 words randomly assigned from the trait word list we developed on all 24 dimensions with the following question: "When you use the word [word name] to describe someone, are you saying they are [low anchor, coded 1], [high anchor, coded 2], or [neither, coded 3]?" Participants rated words on one dimension per survey page, with the 10 words displayed on a screen as three-option multiple-choice questions. On average, each word was rated by 10.78 raters, and all words were rated by at least 10 people who passed our attention check. Participants were also explicitly instructed that "if you do not know the meaning of a word, you can look up the definition."

We calculated the proportions of times each word was rated with either the low anchor or the high anchor (vs. "neither") to quantify the relevance of a word on a specific dimension. A relevance score of 0.70 means that 70% of participants rated a word as communicating the dimension. Figure 3 shows the relevance scores for four words across all dimensions.

How Many Trait Dimensions Are Needed to Cover the English Trait Lexicon?

We first tested whether the 24 dimensions from our review were comprehensive. In other words, were our 2,847 trait words relevant to at least one of the dimensions? Testing this question requires setting a threshold for "relevance." For example, a threshold of 0.20 would include words that at least 20% of raters viewed as relevant to a given dimension, yielding many false positives. By contrast, a conservative relevance score of 1.00 would only include words that all raters viewed as relevant to a particular dimension, yielding false negatives. There was no obvious relevance threshold for this analysis, just like there was no clear threshold for separating trait words from nontrait words in Study 1. For this reason, we calculated results for different relevance thresholds ranging from 0.20 to 0.80.

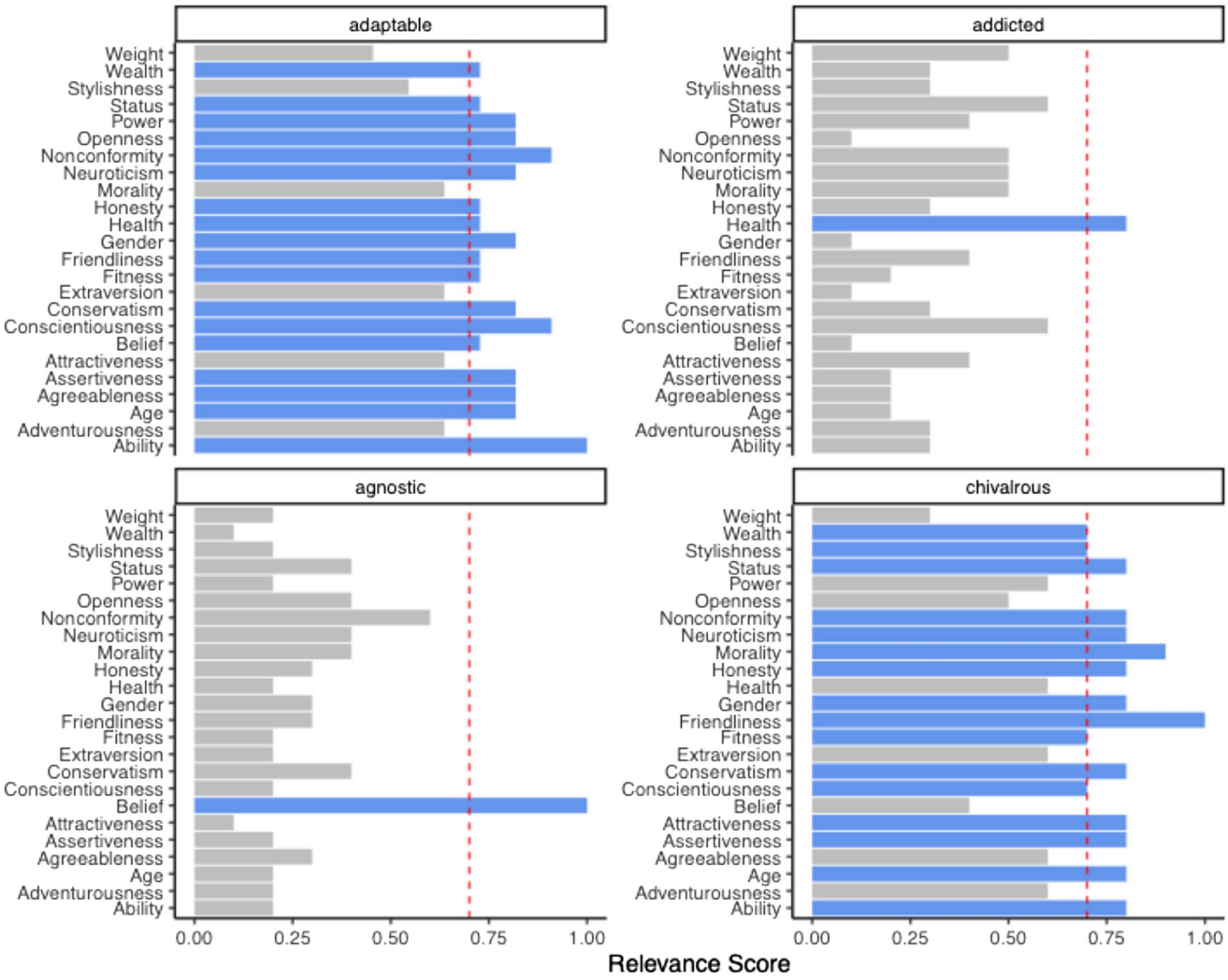
In Figure 4, we display the coverage of each trait dimension at different relevance thresholds. As we applied a stricter threshold, the coverage of dimensions decreased accordingly. But under a wide range of thresholds, the 24 dimensions that we tested were comprehensive, with high relevance to nearly every word in our list. When we used a 0.5 threshold, only three words (0.11%) were not covered by these 24 dimensions. With a 0.7 threshold, 111 words (3.90%) were not covered. In terms of individual dimensions, "friendliness" covered the greatest number of words, whereas "weight" covered the fewest words.

To strike an adequate balance between coverage (i.e., recall) and precision, we set 0.70 as the threshold in subsequent analyses. Although any threshold is a subjective researcher choice, we note

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 3

Four Words Displayed With Their Relevance Scores Across the 24 Dimensions



Note. Blue bars are cases in which a word was judged as relevant to a particular dimension by no less than 70% of the raters (relevance score ≥ 0.7). Red dashed lines represent a relevance threshold = 0.70. For example, the word "agnostic" was judged as most relevant to the dimension of "belief" but also somewhat relevant to "nonconformity." See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

that our findings replicate across a variety of different thresholds (see Supplemental Material 3). We ultimately chose 0.70 because it allowed us to analyze almost all of our words while ensuring that a majority of raters deemed words as relevant to their respective dimensions. When we visually inspected the 111 "uncovered" words with a 0.70 threshold, we did not find any evidence of a single "omitted" dimension. Instead, most of the uncovered words were rare or had noisy ratings.[2]

Is There a Parsimonious Structure of These Dimensions?

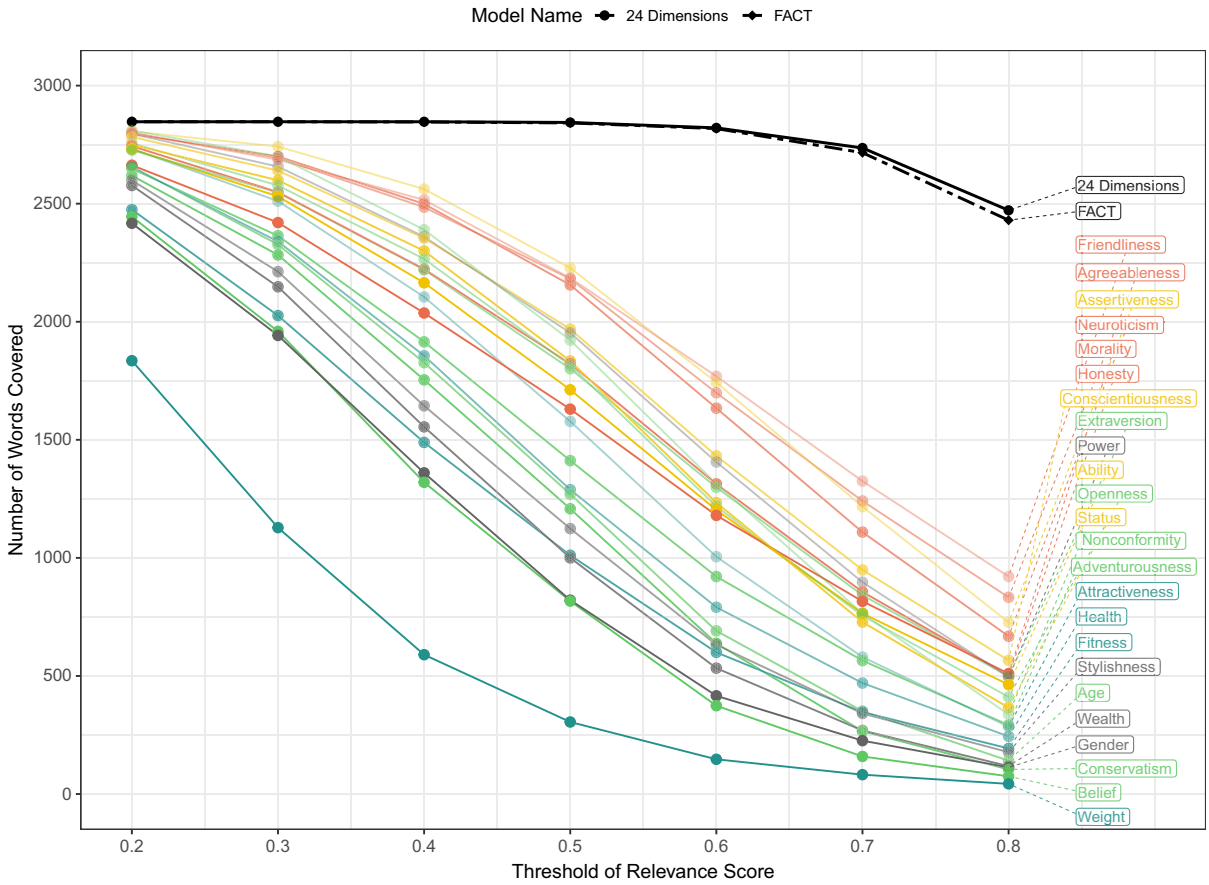
We next tested whether and how the 24 dimensions could be reduced to a lower dimension structure. For example, we reasoned that many words expressing "friendliness" would also express "agreeableness," which would suggest that a single latent factor underlies both dimensions. These coexpressions were common. Panel A of Figure 5 is a matrix illustrating the correlation of all dimension pairs. Cells with high correlations are dimensions that are often coexpressed by the same words.

We used an EFA as our dimension reduction strategy. We fit this EFA to a 2,847-word × 24-dimension matrix in which each cell represented the perceived relevance of a given word to a specific trait dimension. Our factor analysis used principal axis factoring, oblimin rotation, and parallel analysis to determine the number of factors.

[2] We found that many of them are rare words that ordinary native speakers may not fully understand but that nonetheless show relevance to our 24 dimensions, such as "wanton" (reckless, cruel, promiscuous, unrestrained), "avuncular" (suggestive of an uncle especially in kindliness or geniality), and "quixotic" (unrealistically idealistic). A few others were not covered because of noise in human data, such as "deftness," which scored 0.69 on the ability dimension.

Figure 4

Coverage of Trait Dimensions Across Relevance Thresholds



Note. Relevance refers to the proportion of raters who judged a word as communicating a given trait dimension, based on the Study 2 rating task. A threshold of .70 means that a word is counted as covered by a dimension only if at least 70% of raters judged the word as relevant to that dimension. As the relevance threshold increases, the precision of words included in each dimension increases, but the number of words covered by each dimension decreases. Colored dots and lines show coverage for each of the 24 candidate dimensions. Black dots and lines show coverage for all 24 dimensions combined and for the four-factor FACT structure (see the next section). For instance, the weight dimension included about 1,800 words at a threshold of .20 but only 82 words at a threshold of .70. FACT = fitness, agency, communion, and traditionalism. See the online article for the color version of this figure.

Oblimin rotation allows factors to correlate, reflecting the assumption that human traits are often interrelated. Parallel analysis provides a principled method for determining the number of factors in an EFA, as it compares observed eigenvalues to those obtained from random data with the same dimensions, helping to identify factors that explain more variance than expected by chance. We also assessed the robustness of the solution with several supplementary analyses. In Supplemental Material 11, we used a word-stratified bootstrapped EFA (i.e., for each word, repeatedly resampling participant ratings with replacement to reconstruct the word-by-dimension relevance matrix and reestimate the factor structure); in Supplemental Material 14, we reestimated the model after removing highly polysemous words; in Supplemental Material 15, we repeated the EFA on a stricter, human-anchored subset of words; in Supplemental Material 17, we repeated the EFA after trimming lower probability words from the final lexicon; and in Supplemental Material 18, we evaluated the stability of the word-level rating profiles under a conventional interrater-reliability framework by calculating intraclass correlation coefficient (ICC) estimates—absolute-agreement ICC(2, *k*) and consistency ICC(3, *k*)—for each word. We replicated the main results reported below across these robustness checks.

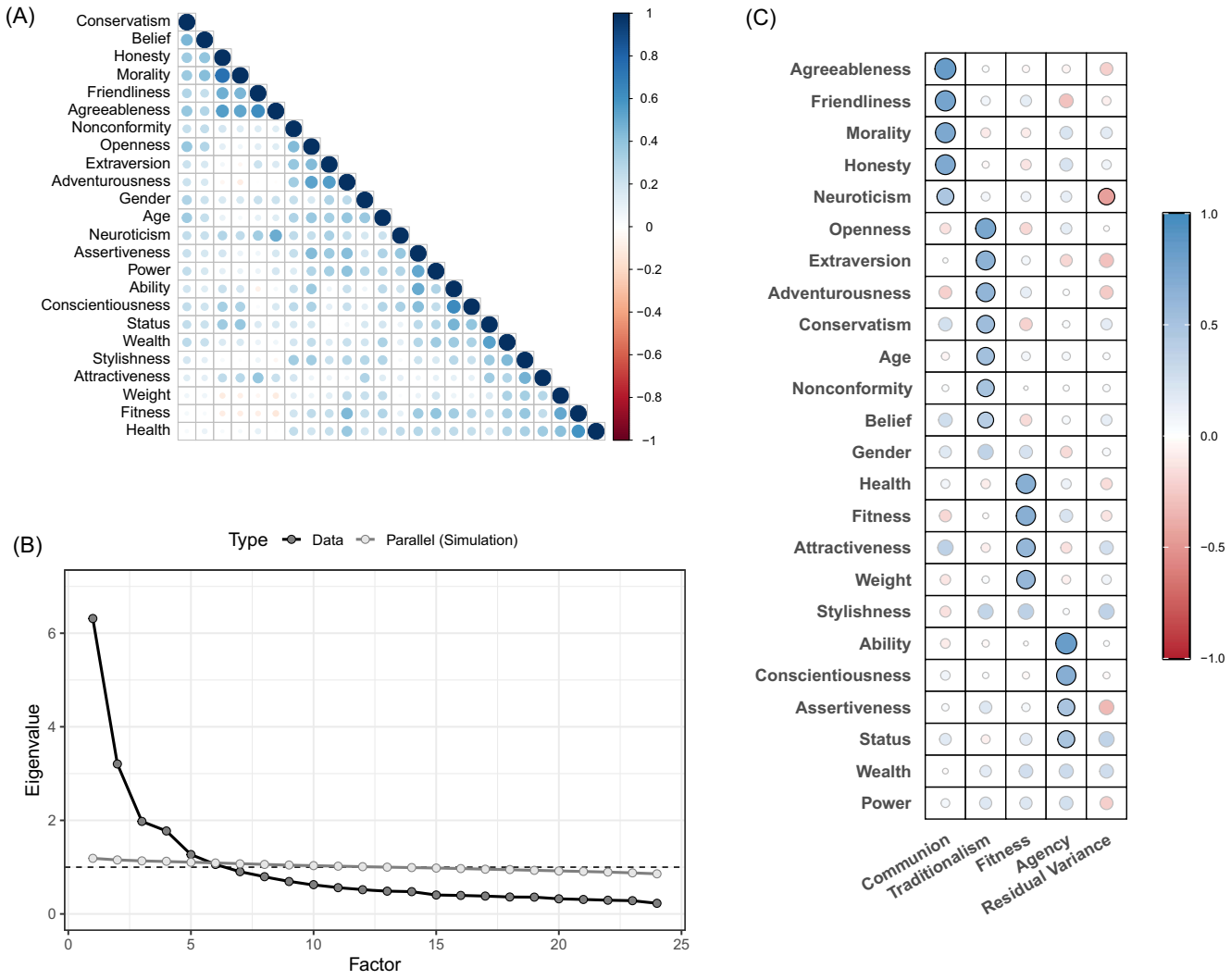
We obtained a five-factor solution (see Figure 5, Panel B), which cumulatively accounted for 45.91% of the total variance. However, closer inspection suggested that only four of these factors were substantively meaningful. The fifth factor accounted for 4.54% of the variance but contained no dimension with a primary loading above .40 (neuroticism cross-loaded on the first and fifth factors). We called this fifth factor "residual variance." The remaining four factors all had meaningful primary loadings. The four substantively interpretable rotated factors accounted for 12.54% (communion), 11.19% (traditionalism), 8.81% (fitness), and 8.82% (agency) of the variance, respectively.

Communion. The first factor accounted for 12.54% of the variance and included agreeableness, friendliness, morality, honesty,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 5

Results of Exploratory Factor Analysis for Human Relevance Ratings



Note. (A) The correlations among dimensions on relevance scores. (B) The scree plot and the result of parallel analysis. (C) The exploratory factor analysis result. The radii of circles represent the absolute value of the factor loadings. The absolute values over .40 are highlighted. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

and neuroticism, which cover the two facets of the horizontal dimension in Abele et al.'s (2021) integrated framework, or the communion dimension in Koch et al.'s (2016) model and the communion dimension in Abele and Wojciszke's (2007) model. We therefore called this factor "communion."

Traditionalism. The second factor accounted for 11.19% of the variance and consisted of openness, extraversion, adventurousness, conservatism, age, nonconformity, and belief. All of these dimensions are relevant to a person's progressive versus conservative orientation, which has been summarized as a general tendency to exploit familiar, socially sanctioned options (more conservative) versus explore novel ones in their social and material environments (more progressive; Koch et al., 2025). We named it "traditionalism."

Fitness. The third factor accounted for 8.81% of the variance and was composed of health, fitness, attractiveness, and weight, which describe a person's physical fitness. These features have also been covered in some recent studies in person perception (Hester & Gray, 2018; Hester & Hehman, 2023; Wilson et al., 2017). We named it "fitness."

Agency. The fourth factor accounted for 8.82% of the variance and included ability, conscientiousness, assertiveness, and status, which roughly correspond to the two facets of the vertical dimension in Abele et al.'s (2021) integrated framework, or the agency dimension in Koch et al.'s (2016) model. We thus continue to label this factor "agency."

Remaining Dimensions. There were four dimensions remaining that did not load over 0.40 on any of the factors. These were gender, stylishness, wealth, and power. We labeled these dimensions "uncategorized."

Panel C of Figure 5 displays the loadings of all dimensions on all factors with a matrix wherein dimensions are organized as rows and

factors are organized as the columns. We collectively refer to these four meaningful factors as FACT. In total, the FACT model covered 95.40% (i.e., 2,716 words) of all 2,847 trait words at the threshold of 0.7.

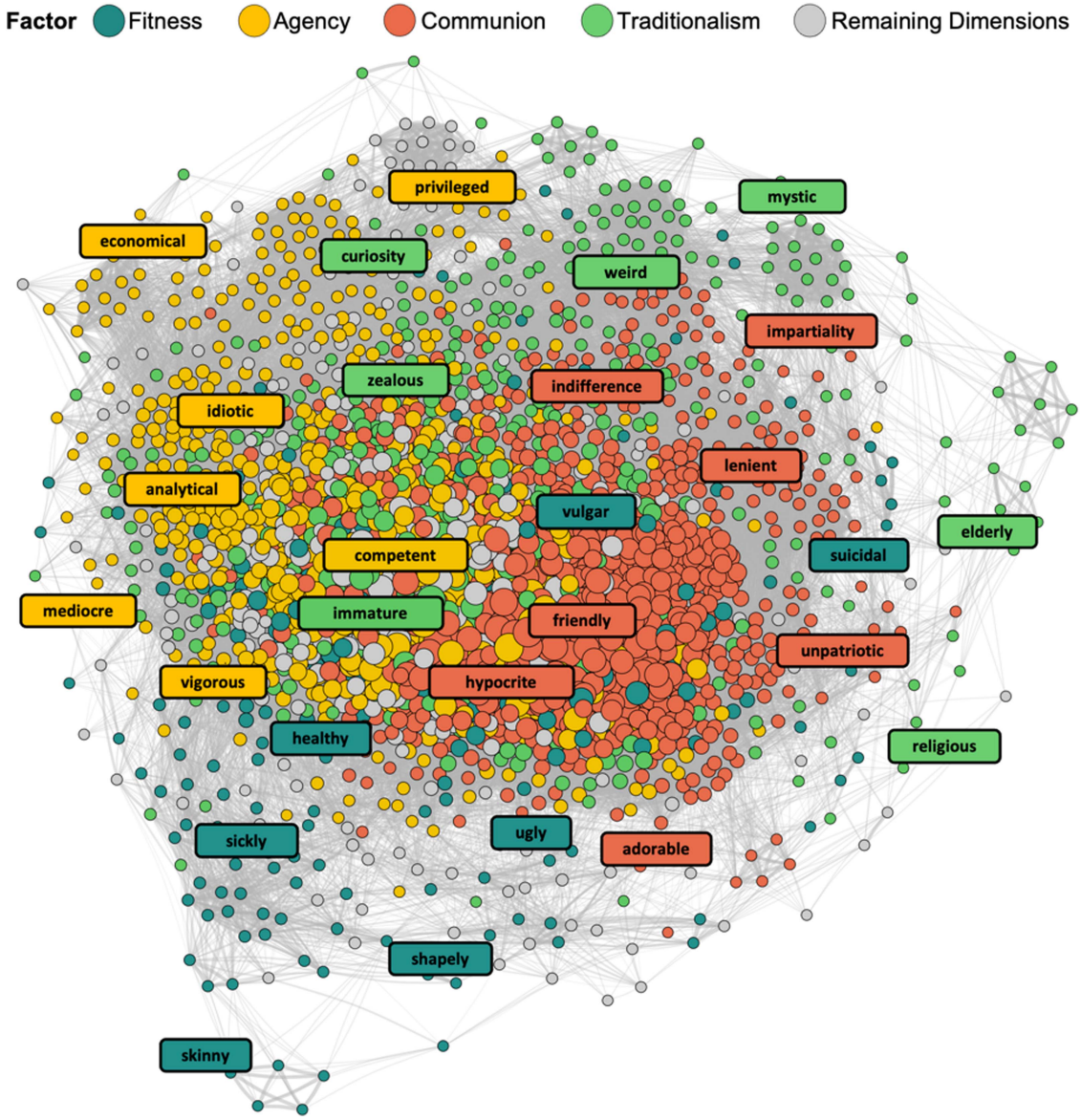
How Coherent Is the Trait Space?

Our final analyses focused on the semantic coherence of trait space. For example, were "communion" words about honesty and morality more clustered together than "agency" words about ability and assertiveness?

Initial exploration revealed that coherence was generally low for all of these trait factors. To show this visually, Figure 6 plots the nomological network of trait words, with edges weighted by Jaccard similarity: the number of trait dimensions expressed (above a 0.70 threshold) by both words in a given word pair divided by the total number of dimensions expressed by either of the words in the pair.

Figure 6

A Network of Trait Words Based on Shared Dimensions



Note. Each node represents a separate word. Edge weights represent the Jaccard similarity of trait words based on underlying trait dimensions. For example, a Jaccard similarity of 1.00 would mean that Trait Word A and Trait Word B express exactly the same dimensions. Node color represents the fitness, agency, communion, and traditionalism dimension with the highest relevance. Node sizes represent the number of dimensions a word expresses; the larger the node, the more trait dimensions the word expresses. An interactive version of this visualization is available online (https://societyinmind.net/assets/trait-network/). See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Each word is shaded by its most relevant FACT factor. Although words clearly cluster by these FACT factors, this clustering is not very modular. In particular, words related to fitness and traditionalism are scattered across the entire network rather than tightly clustered in specific locations.

We measured coherence across the FACT factors by calculating generality scores, defined as the degree that words express many dimensions at once. For example, in Figure 3, the word "adaptable" is generalized because raters judged it as relevant to many dimensions, whereas the word "addicted" is specific because raters judged it as relevant to just one dimension. We reasoned that a given factor would be coherent if the words in this factor applied broadly to other dimensions within the same FACT factor (high within-factor generality) but did not spill over into the other FACT factors (low between-factor generality).

One complication with calculating within-factor generality was that some dimensions—like friendliness and agreeableness—were highly conceptually similar, and words expressing both of these dimensions might not necessarily reflect high within-factor generality. To correct for this complication, we calculated a redundancy coefficient φ for each dimension by calculating its average correlation with other dimensions in the same factor and weighted redundant dimensions less than nonredundant dimensions using the weighting formula $1/(1 + \varphi)$. We calculated between-factor generality as the proportion of FACT factors that each word was relevant to, with no weighting scheme. For example, the word "intelligent" was judged as relevant to one dimension from agency (ability, relevance = 0.91) and one dimension from traditionalism (openness, relevance = 0.82), giving it a between-factor generality score of 0.5 (i.e., 2/4). To ensure that the results did not depend on this particular redundancy-weighting scheme, we also report robustness checks using alternative treatments of redundancy in Supplemental Material 16.

When we estimated the average levels of within- and between-factor generality for each dimension (see Figure 7), we found that communion dimensions had the highest within-factor generality ($M = 0.66$, $SE = 0.02$) and the lowest between-factor generality ($M = 0.69$, $SE = 0.01$), suggesting that they were the most coherent. Agency dimensions had the second highest within-factor generality ($M = 0.64$, $SE = 0.03$) and second lowest between-factor generality ($M = 0.73$, $SE = 0.01$). Traditionalism and fitness dimensions were less coherent, with lower rates of within-factor generality (Ms = 0.42 and 0.52, SEs = 0.01 and 0.06) and higher rates of between-factor generality (Ms = 0.76 and 0.76, SEs = 0.01 and 0.05).

These coherence analyses are consistent with what we see in the full nomological network. Fitness and traditionalism words appear to be the brokers of this network, projecting into clusters of agency and communion words. By contrast, agency and communion appear to be more central to the network, clustering more tightly together, with communion words forming the most coherent cluster.

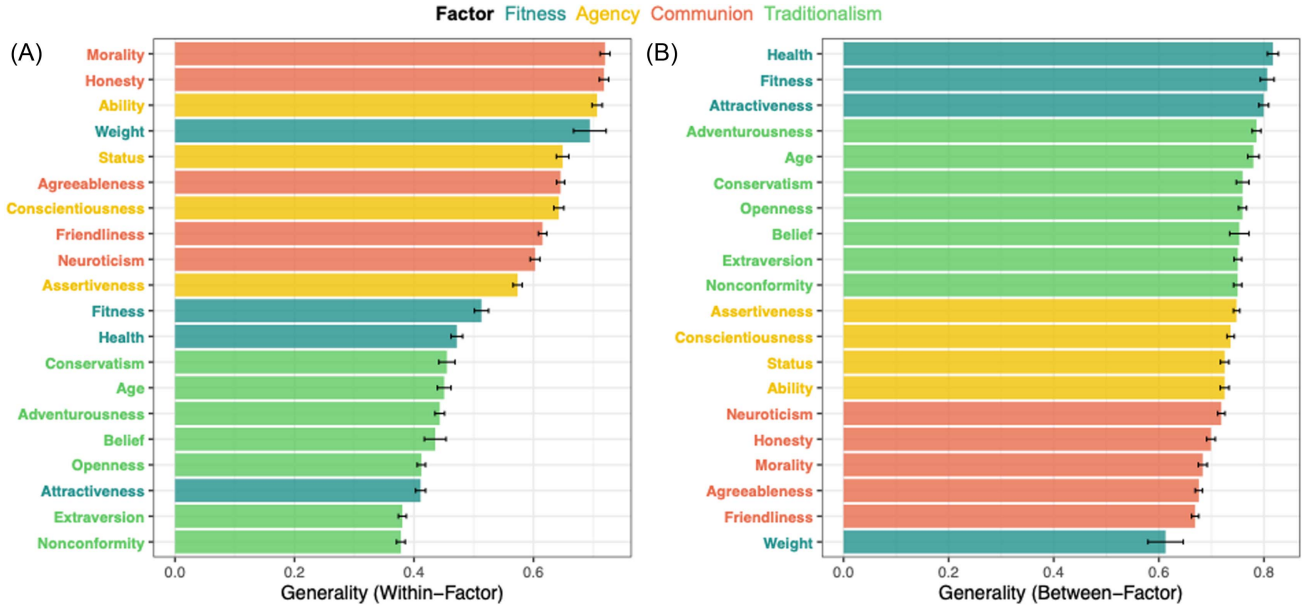
Discussion

Using a comprehensive trait list, Study 2 revisited three important questions about human traits: What would a comprehensive dimensional structure of human traits look like? What would a parsimonious structure look like? How coherent is the trait space?

The 24 dimensions that we selected for our study are good examples of a high-dimensional structure that covers the full trait space. These dimensions covered 96.10% of our 2,847-word trait list with a relevance threshold of 0.70. In Supplemental Material 4, we considered potential redundancies among these dimensions and

Figure 7

Two Types of Generality Scores at Dimensional Level



Note. (A) The average within-factor generality scores of each dimension. Error bars are standard errors. (B) The average between-factor generality scores of each dimension. Error bars are standard errors. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

systematically removed the most redundant ones using a principled procedure. Even after this reduction, we found that approximately 9–11 dimensions were still needed to cover over 90% of trait words and 15–16 dimensions to cover over 95% of trait words, suggesting that trait space represents a richly structured, high-dimensional network.

We do not attach great importance to specific numbers such as 9–11 or 15–16. These counts are partly an artifact of using our 24 starting dimensions and the particular redundancy-reduction procedure. Accordingly, we emphasize that the specific set of 24 dimensions we sampled was inevitably subjective and was intentionally designed to include some redundancy to maximize lexical coverage; it should therefore not be interpreted as an optimal or definitive taxonomy of trait language. At the same time, we expect the broader conclusion of high dimensionality to be robust to the choice of starting dimensions. Future work could begin with alternative frameworks (e.g., dark-trait factors; Feher & Vernon, 2021), and we anticipate that such choices are unlikely to eliminate the need for a comparatively high-dimensional account of trait language.

With respect to a low-dimensional structure, our EFA recovered a four-factor FACT model consisting of fitness, agency, communion, and traditionalism. The FACT model included dimensions resembling previous low-dimensional models but did not exactly reproduce any specific model. For example, three factors in the FACT model resemble the agency, communion, and beliefs dimensions of the ABC model. The fourth factor, fitness—comprising health, fitness, attractiveness, and weight—has not been a prominent part of previous models, but it may be a critical supplement. It represents a primal factor in person perception that reflects objective physical characteristics. These physical characteristics are highly accessible in social evaluations and play a key role in the formation of first impressions (Freeman & Lin, 2025; Hester & Hehman, 2023). Its relative independence from other factors indicates that fitness is not a direct signal of communion, traditionalism, or competence.

Not all dimensions, however, loaded cleanly onto one of the four broader factors. In particular, gender, stylishness, wealth, and power remained outside the main FACT structure. We do not interpret these dimensions as unimportant or merely residual. Rather, one possibility is that they are especially cross-cutting or context dependent. For example, stylishness and dress can simultaneously signal identity, status, and aesthetic preference (Hester & Hehman, 2023), whereas wealth and power often shape person perception through multiple, partially overlapping routes tied to status, competence, dominance, and group position (Bjornsdottir & Rule, 2017; Oh et al., 2020; X. Wang et al., 2018). Gender may likewise structure person perception through intertwined category, status, and role-based inferences rather than as a marker of a single broader domain (Ridgeway, 2001). From this perspective, their failure to load cleanly onto a single broader factor may itself be informative, suggesting that some parts of trait space are organized less around one dominant axis and more around their ability to enter many different forms of social inference.

Our final analyses focused on the coherence of the FACT factors, which we computed using measures of within- and between-factor generality. The high within-factor generality of communion may reflect that people view different aspects of communion as relatively interchangeable. For example, we may often infer that friendly people are also moral. The within-factor generality of agency was similarly high. By contrast, traditionalism and fitness had lower within-factor generality, suggesting that people may view dimensions within these factors as more distinct. For example, an introverted person can be either open-minded or close-minded, progressive or conservative.

For between-factor generality, we found that fitness and traditionalism were perceived as more generalizable than communion and agency. Fitness traits are usually observable, direct representations of people's physical features, which are highly accessible in impression forming. Neuberg et al. (2020) have described fitness traits as powerful for being "readily perceived," "relatively invariant across time and locations," and "difficult to disguise or fake." As such, fitness traits serve as versatile cues. For instance, fitness can serve as a signal of communion (e.g., physically attractive individuals are often perceived as more trustworthy), traditionalism (e.g., body weight or physique cues are used to identify a fellow "outdoor adventurer"), or agency (e.g., healthy and strong individuals may be perceived as more competent or dominant).

The overall high between-factor generality of traditionalism is also intriguing and resonates with previous findings that ideology traits can be cues of warmth and competence (Koch et al., 2018, 2020). This pattern might arise because traditionalism traits can serve as convenient signals of social identities, and the process of social identification will in turn facilitate group-based inferences of communion and agency (Brewer, 2007; McPherson et al., 2001). One real-world example is affective polarization, where opposing partisans doubt each other's trustworthiness, morality, intelligence, and ability (Dias & Lelkes, 2022; Iyengar & Westwood, 2015; Lelkes, 2016).

Study 3

We next turned to the valence[3] of trait words in Study 3. First, we analyzed whether the valence of trait words was correlated with words' location on the "semantic direction"—from one pole to the other pole—of each dimension and whether dimensions from different FACT factors showed similar correlations. We next tested whether all trait dimensions contained more negative than positive trait words or whether there was variation across dimensions in the number of negative versus positive trait words. One theoretical mechanism for a greater number of negative trait words is that negative information is more heterogeneous, and we therefore require many trait words to communicate many different forms of negativity. We independently tested this mechanism by estimating whether negative (vs. positive) trait words were more semantically specific, communicating a narrower range of dimensions.

Method

Measures

Valence. We extracted valence scores from the National Research Council Canada valence, arousal, and dominance Lexicon v2 (Mohammad, 2025), where each word was rated by multiple native English speakers on a 7-point scale ranging from −3 (*very negative feelings*) to +3 (*very positive feelings*), later linearly transformed to the range −1 to +1, with 0 representing a neutral state. Each word had, on average, about eight valid ratings provided

[3] In our preregistration, we also mentioned an analysis of arousal. However, to streamline the article we report those analyses in Supplemental Material 9.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

by native English speakers, mostly from the United States, with high split-half reliability ($r > .95$). We matched these ratings to our data set by word, resulting in valence scores for 2,402 words ($M = -0.07$, $SD = 0.63$). In some analyses, we created a three-level categorization of words as negative (valence score < -0.3), neutral ($-0.3 \leq$ valence score ≤ 0.3), and positive (valence score > 0.3).

Procedure and Results

How Is Valence Distributed Within Each Dimension?

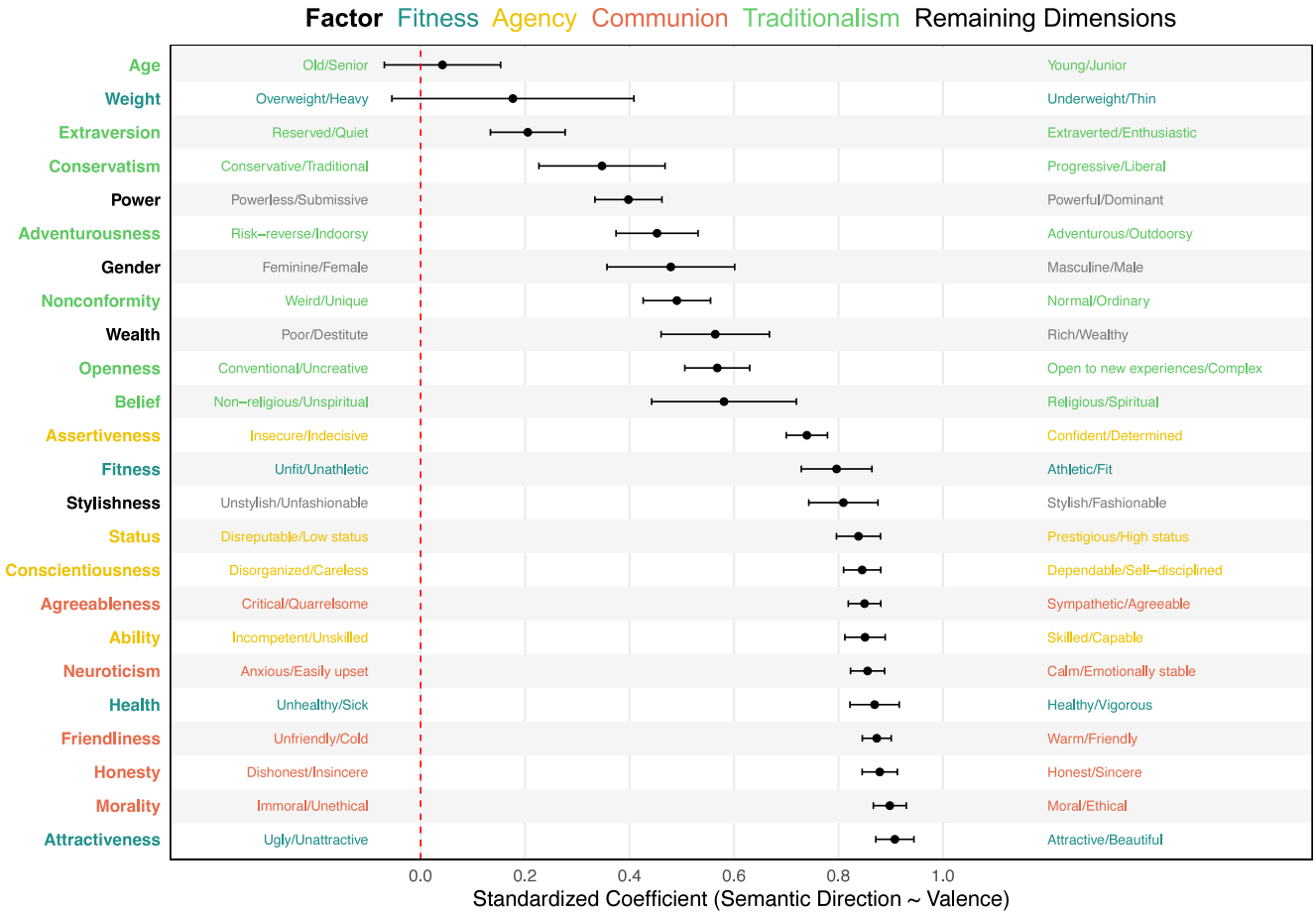
We calculated semantic direction of each word on each dimension using the bipolar scale that we provided raters in Study 2, which is defined as the ratio of the number of "right anchor" ratings to the total number of "left anchor" and "right anchor" ratings. For example, a semantic direction score of 0.90 for agreeableness would mean that 90% of raters evaluated the word as meaning sympathetic/agreeable and the remaining 10% of raters evaluated the word as meaning critical/quarrelsome. Each semantic direction score was calculated based on at least seven raters because we adopted a relevance threshold of 0.70 in our analyses. We classified words as left anchored (score < 0.33), neutral ($0.33 \leq$ score ≤ 0.67), and right anchored (score > 0.67).

We then calculated how the valence of each word correlated with the semantic direction of each word. We computed this correlation for each trait dimension, with stronger correlations signifying that a trait dimension had oppositely valenced poles. For example, a correlation of .90 for agreeableness would mean that being rated as higher in sympathetic/agreeable corresponded very strongly to being rated as more positive. To facilitate comparisons, we arranged the semantic directions of some dimensions to make all correlations positive, considering that the left–right positioning of anchors was arbitrary (e.g., we could have placed "young" and "old" on either side of the scale).

As shown in Figure 8, attractiveness had the best alignment between semantic direction and valence, $r(496) = .91$, $p < .001$, meaning raters consistently rated "attractive/beautiful" as more positive than "ugly/unattractive." Health also exhibited a high correlation, $r(412) = .87$, $p < .001$, meaning raters consistently preferred "healthy/vigorous" to "unhealthy/sick." Otherwise, however, the most aligned dimensions belonged to the communion factor, followed by the agency factor. The agency dimension with the weakest

Figure 8

Associations Between Semantic Direction and Valence for Each Dimension



Note. Standardized coefficients in models regressing semantic direction on valence (equal to correlation coefficients here). Anchors are arranged on the left side versus right side so that all correlations are positive. Error bars are 95% confidence intervals. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

correlation was assertiveness, $r(1066) = .74$, $p < .001$. Assertiveness represented a tipping point, after which the remaining 11 dimensions had weaker correlations.

All but four (i.e., weight, power, gender, and wealth) of these weaker correlations involved traditionalism dimensions, and all traditionalism dimensions came after the assertiveness tipping point. For example, words signifying that someone was progressive/liberal were viewed as moderately more positive than words signifying that someone was conservative/traditional, $r(226) = .35$, $p < .001$, but even this moderate correlation may be partly attributed to the fact that Prolific workers are mostly liberal (Douglas et al., 2023). Weight, $r(69) = .18$, $p = .140$, and age, $r(307) = .04$, $p = .462$, were the only two dimensions whose semantic direction did not correlate significantly with valence.

When we aggregated these correlations to the level of the FACT factors, communion exhibited the highest average correlation between valence and semantic direction, $M_r = 0.87$, $SE = 0.01$, followed by agency, $M_r = 0.82$, $SE = 0.03$, and fitness, $M_r = 0.69$, $SE = 0.17$. Traditionalism showed the lowest average correlation, $M_r = 0.38$, $SE = 0.08$.

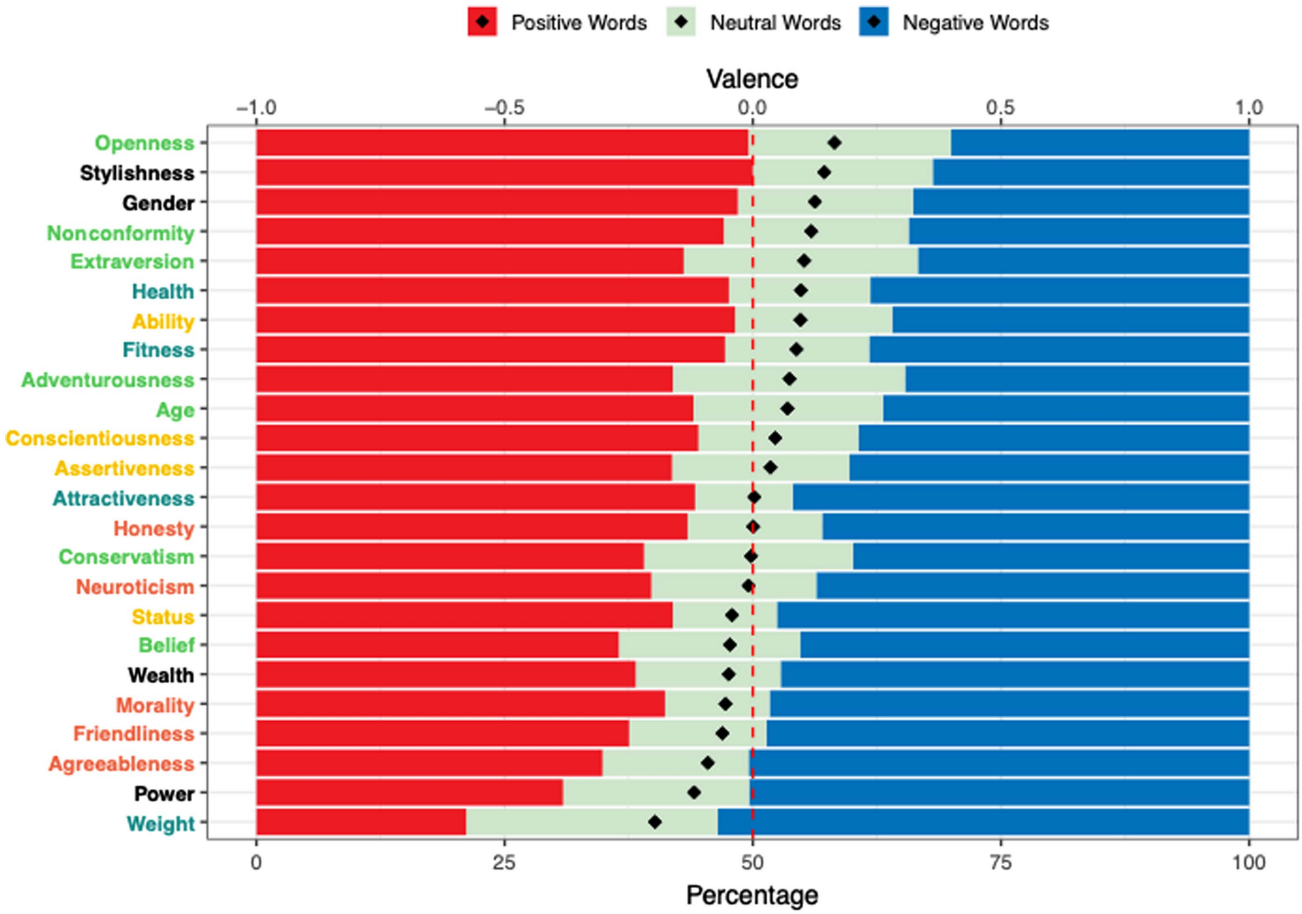
Are There More Negative or Positive Trait Words?

We summarized the number of positive and negative trait words for each dimension and ranked them by the average valence scores. We found that 13 dimensions had more positive trait words than negative trait words, while the remaining 11 dimensions had more negative words than positive words. Figure 9 displays the prevalence scores for each dimension. At the factor level, communion had the lowest valence, $M_{valence} = -0.07$, $SE = 0.01$, followed by agency, $M_{valence} = -0.04$, $SE = 0.01$; fitness, $M_{valence} = 0.00$, $SE = 0.02$; and traditionalism, which had the highest valence, $M_{valence} = 0.04$, $SE = 0.01$. These results supported the hypothesis that the ratio of negative to positive trait words would vary systematically across trait dimensions.

One of the initial reasons we thought that there may be more negative words than positive words was that negative words are more specific, and hence more semantically differentiated, than positive words (Alves et al., 2017). Interestingly, we did see evidence for this prediction, even though there were not uniformly more negative words than positive words. In all but four dimensions,

Figure 9

Valence Distribution for Each Dimension



Note. Bars show the percentages of positive, neutral, and negative words within each dimension, ordered from left to right for visual comparison. The lower axis represents percentages, whereas black diamonds indicate dimension-level average valence scores on the upper axis. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

higher valence (i.e., positivity) correlated with within-factor generality. Among the four dimensions with nonsignificant results, three (i.e., weight, fitness, health) were from the fitness factor. Positive valence also correlated with more between-factor generality in all but five dimensions. Among these exceptions, four (i.e., conservatism, extraversion, adventurousness, and belief) were from the traditionalism factor. These results are displayed in Figure 10.

Discussion

Study 3 systematically analyzed the valence of the trait space, focusing on two main questions. Our first question concerned the distribution of valence within trait dimensions. We found that the valence of words in the communion, agency, and fitness factors (except weight) aligned strongly to their semantic direction (with correlations all higher than .70). This suggests that our human raters strongly agreed on which ends of these dimensions were positive versus negative. On the other hand, semantic direction was weakly correlated with valence for traditionalism dimensions such as conservatism and extraversion, indicating substantially less agreement.

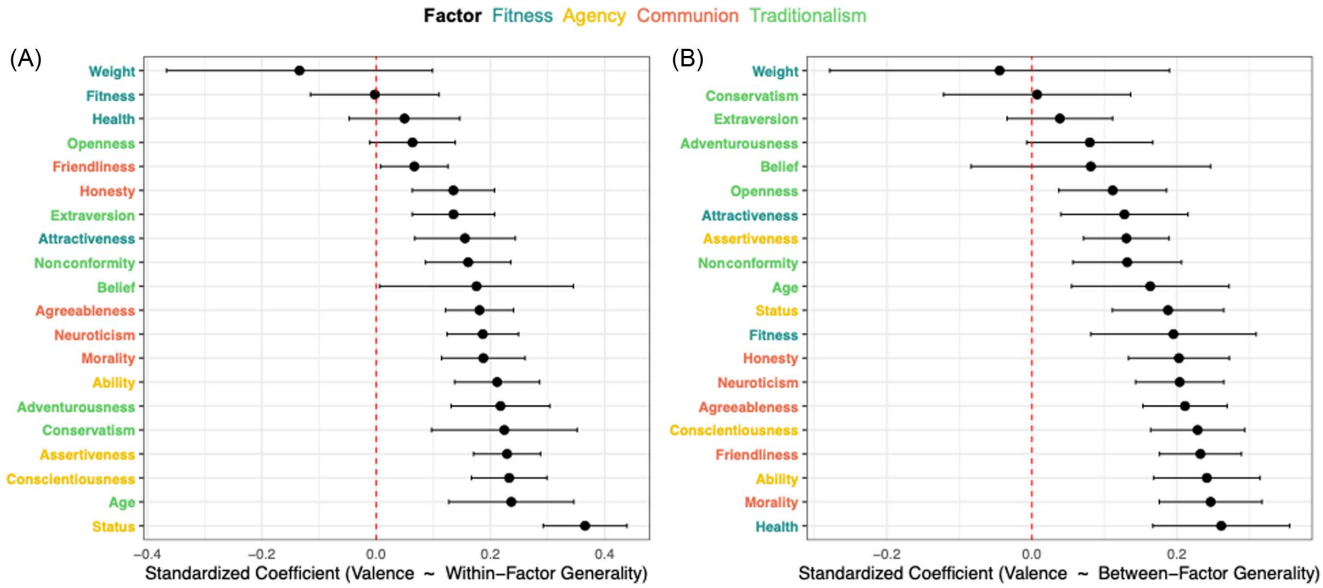
Theoretically, these results suggest that some trait dimensions may map closely to basic human needs, whereas others may reflect personal or culturally shaped preferences. On the one hand, the positivity of communion may reflect people's desire for intimacy and relatedness; the positivity of agency may reflect people's desire for power, autonomy, and competence; and the positivity of fitness may reflect fundamental biological needs for health, physical integrity, and reproductive fitness. On the other hand, the mixed valence of traditionalism may reflect the fact that people's preferences for traditionalism are rooted in their personal and cultural background (Graham et al., 2009; Hofstede et al., 1990; Leung & Bond, 2004). Even the moderate positivity of progressive/liberal (compared with conservative/traditional) words may have been an artifact of the fact that the Prolific sample largely comprised English-speaking, liberal-leaning individuals (Douglas et al., 2023). In a different sample, this correlation may have disappeared or reversed. In the General Discussion section, we return to this point and propose a unifying framework, grounded in game theory (Colman, 2003), to explain when valence and semantic direction should be tightly coupled and when they may diverge.

Our second question concerned whether there should be more negative than positive trait words in the English lexicon. Prior theories offer mixed perspectives on this question. For example, one perspective suggests that there might be more negative words than positive words because negativity is more specific than positivity (Alves et al., 2017), and we need different words to describe the many flavors of negativity. We did indeed find that negative trait words were more specific than positive trait words for most dimensions—communicating a narrower set of dimensions. However, we did not find that negative trait words were uniformly more numerous than positive trait words. Instead, the relative number of positive and negative words varied widely across dimensions.

For communion, there were more negative words than positive words, but the opposite was true for agency, traditionalism, and fitness, although the exact ratio varied widely across specific dimensions. The variation across dimensions may be because people develop words to communicate diagnostic information, and different information is diagnostic for different dimensions (Skowronski & Carlston, 1987, 1989). For example, because most people are friendly and honest, negative words about communion are diagnostic, whereas the opposite is true for agency. A separate theory of "error

Figure 10

The Relationship Between Valence and Two Types of Generality



Note. (A) The standardized regression coefficients using valence to predict within-factor generality of each dimension—error bars are 95% confidence intervals. (B) The standardized regression coefficients using valence to predict between-factor generality of each dimension—error bars are 95% confidence intervals. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

management" (Haselton & Buss, 2000; Johnson et al., 2013) emphasizes that people communicate more words in contexts wherein prediction errors are costly. To this point, theories have argued that false positives about communion (i.e., mistaking dishonest for honest) are more costly than false negatives, and people should therefore create more negative words about communion. However, the opposite is true for agency. Both a diagnosticity and an error management perspective predict the dimension-specific ratios of positive to negative words that we found.

It is important to note that we can offer the above interpretations only because, for communion and agency dimensions, semantic direction closely aligns with valence: Positively valenced words tend to represent the presence or abundance of the target traits, while negatively valenced words describe their absence or deficiency. However, this logic does not apply cleanly to traditionalism dimensions, where we found semantic direction and valence did not align well—for instance, both "traditional" and "progressive" can carry positive connotations. Nonetheless, we did observe the highest average valence in traditionalism dimensions. This may be because, unlike the high interpersonal costs of negative traits in communion (e.g., unkindness) or agency (e.g., incompetence), undesirable preferences in traditionalism are often less disruptive. It is not inherently aversive to encounter someone who holds different lifestyle values, even if we disagree with them. Therefore, people may not form strong negative sensitivity to them (Baumeister et al., 2001). Instead, traditionalism traits may be more commonly used for self-description, serving identity-affirming or group-signaling purposes, and thus are more likely to be framed in a positive light (Brewer, 1991; Greenwald et al., 2002; Steele, 1988).

Study 4

In our final study, we analyzed the evolution of the English trait lexicon. We wanted to know whether the expansion of this trait lexicon has encompassed major changes to trait space or whether trait space has remained relatively stable over the last several centuries. Specifically, we tested (a) whether the same trait dimensions have always had more versus fewer words (i.e., are dimensions consistent in their scope), (b) whether words from some trait dimensions have always been used more frequently in written texts than words from other trait dimensions (i.e., are dimensions consistent in their frequency), and (c) whether the semantic structure of traits has changed over time or whether it has been historically stable (i.e., are dimensions consistent in their interrelationships).

Method

Measures

Origination Dates. We used words' first attestation date recorded by the Oxford English Dictionary as a proxy of origination date (Williams, 2024). These attestation dates are not necessarily the same as the birth dates of words because it takes time for a spoken word to be written down, and there may be written instances of some words earlier than the Oxford English Dictionary's records. That said, they are still reasonable estimates. In the original data, some words had several first attestation dates because their respective parts of speech were first used at different times. In these cases, we chose the earliest date among them as the origination date of the word.

Frequency of Use. We extracted the yearly frequency of words in our list from 1800 to 2000 from the English Fiction corpus of Google Books database (Y. Lin et al., 2012). We chose this specific corpus because it has stable content over time, unlike the Ngram corpus (Pechenick et al., 2015). The consistent nature of the English Fiction corpus lends us more confidence that changing word frequency reflects changes in the usage of concepts rather than mere changes in corpus genre composition.

Historical Trait Word Embeddings. To investigate the historical structure of trait words in the English language, we used the pretrained HistWords embeddings of English words from 1800 to 1999 (Hamilton et al., 2018). These are a set of word2vec embeddings of English words in a 300-dimensional space, which are trained on the Google N-Grams eng-all corpus, and organized by decade. Therefore, we got 20 decadal word embedding models in total. We chose these models because they were particularly trained for the purpose of longitudinal comparisons, and they do a better job than other models of addressing the problem of data scarcity in early decades.

Procedure and Results

Stability and Change in the Size and Frequency of Trait Dimensions

Our initial analyses focused on changes in the number of words belonging to each trait dimension (i.e., its size or scope). We began by summarizing the average origination dates of words in each trait dimension. As shown in Panel A of Figure 11, the "youngest" dimension was conservatism, $M_{age} = 1{,}544.46$, $SE_{age} = 11.81$, while the "oldest" dimension was gender, $M_{age} = 1{,}481.84$, $SE_{age} = 15.17$. At the factor level, fitness had the earliest average origination date, $M_{age} = 1{,}492.56$, $SE_{age} = 3.81$, followed by communion, $M_{age} = 1{,}502.29$, $SE_{age} = 6.86$, and agency, $M_{age} = 1{,}511.03$, $SE_{age} = 3.34$. Traditionalism's average origination date was the latest, $M_{age} = 1{,}521.78$, $SE_{age} = 5.27$.

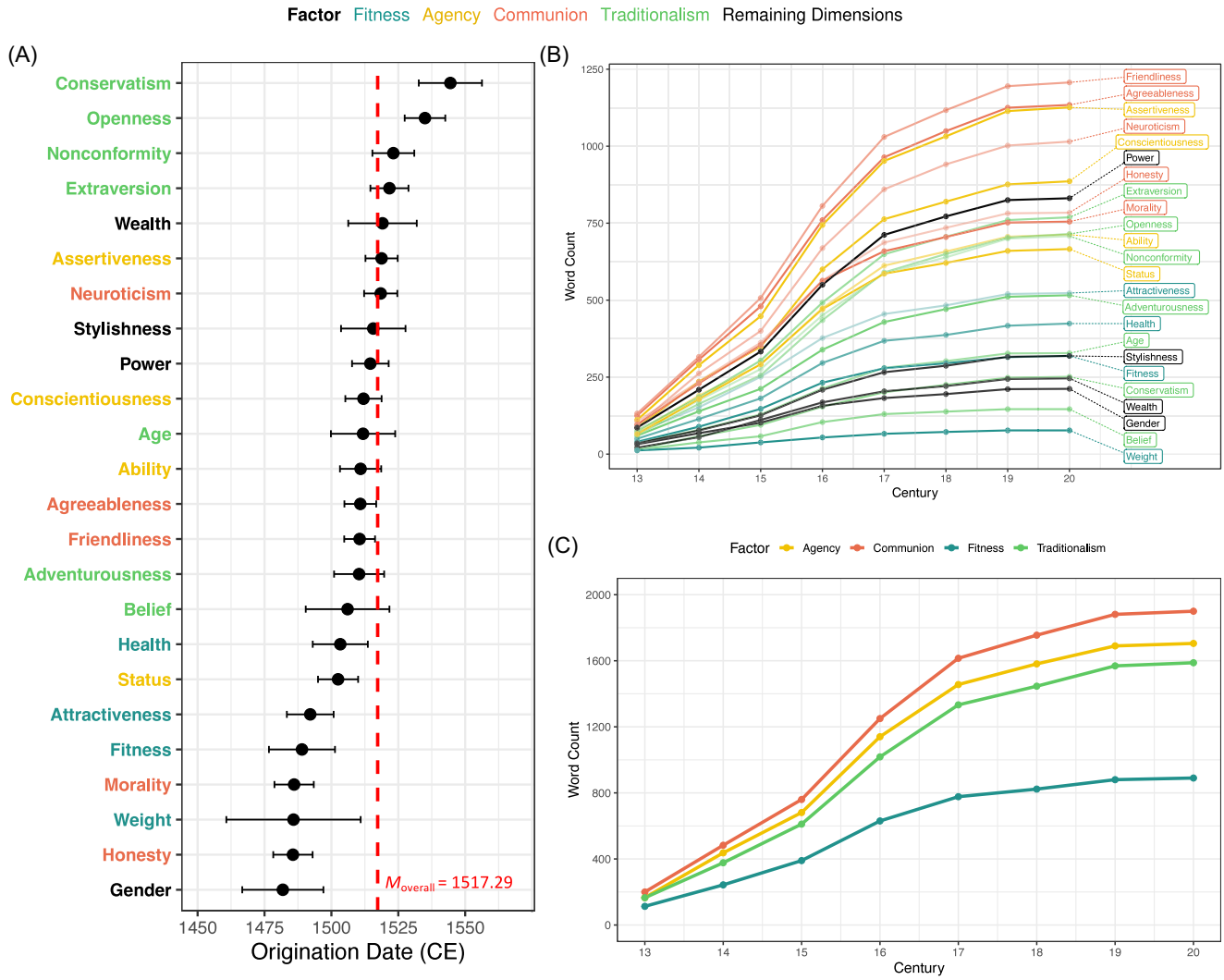
We then charted each dimension's trajectory of vocabulary size and growth rate over time (see Figure 11, Panel B). This analysis revealed remarkable historical stability. For example, there have been more words conveying friendliness than assertiveness since the 13th century. The rank order of the dimensions in the 13th century correlates nearly perfectly with the rank order in the 20th century, $r(22) = .97$, $p < .001$. Panel C of Figure 11 further aggregated the word count to each factor. There have always been more trait words related to communion, followed by agency and traditionalism. Fitness has always had the fewest number of words.

We next analyzed trends in frequency of word usage from 1800 to 2000. As illustrated in Panels A and B of Figure 12, there was also notable stability in these trends: The average frequencies of communion and agency words were consistently highest over time, followed by the average frequency of traditionalism words. These three factors dominated trait language in the 1800–2000 interval that we analyzed. In every one of these years, communion, agency, and traditionalism were each relevant to over 50% of all trait words used in text. By contrast, fitness words were less frequently used. Only

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 11

Origination Dates and Trajectories of Word Count of Each Dimension



Note. (A) Ordered average origination dates of different trait dimensions. Error bars are standard errors; red dashed line is the average origination date of all trait words; (B) The number of words belonging to each dimension over time; (C) the number of words belonging to each factor over time. CE = Common Era. See the online article for the color version of this figure.

about 43% of trait-word usage between 1800 and 2000 was relevant to fitness.

However, there were also significant changes in frequency of use. Panel C of Figure 12 shows linear slopes for each dimension. Specifically, most communion dimensions declined in usage. For example, morality ($b = -5.78 \times 10^{-4}$, $SE = 1.12 \times 10^{-5}$, $p < .001$, $\beta = -.96$, 95% CI [-6.00×10^{-4}, -5.56×10^{-4}]) and honesty ($b = -2.55 \times 10^{-4}$, $SE = 8.62 \times 10^{-6}$, $p < .001$, $\beta = -.90$, 95% CI [-2.72×10^{-4}, -2.38×10^{-4}]) decreased by 18.16% and 10.12%, respectively, between 1800 and 2000. On the other hand, most agency and traditionalism dimensions experienced rising usage. For example, ability ($b = 8.17 \times 10^{-4}$, $SE = 1.64 \times 10^{-5}$, $p < .001$, $\beta = .96$, 95% CI [7.84×10^{-4}, 8.49×10^{-4}]) and openness ($b = 4.46 \times 10^{-4}$, $SE = 1.27 \times 10^{-5}$, $p < .001$, $\beta = .93$, 95% CI [4.21×10^{-4}, 4.71×10^{-4}]) increased by 32.68% and 26.16%, respectively. Among all dimensions, conservatism experienced the largest growth in usage frequency between 1800 and 2000 (45.53%).

Fitness dimensions showed mixed trends. Attractiveness declined significantly ($b = -4.11 \times 10^{-4}$, $SE = 9.18 \times 10^{-6}$, $p < .001$, $\beta = -.95$, 95% CI [-4.29×10^{-4}, -3.93×10^{-4}]), decreasing by 23.70% between 1800 and 2000. At the factor level, frequency increased for traditionalism (22.81%) and agency (16.74%) but decreased for communion (−7.96%) and fitness (−1.78%).

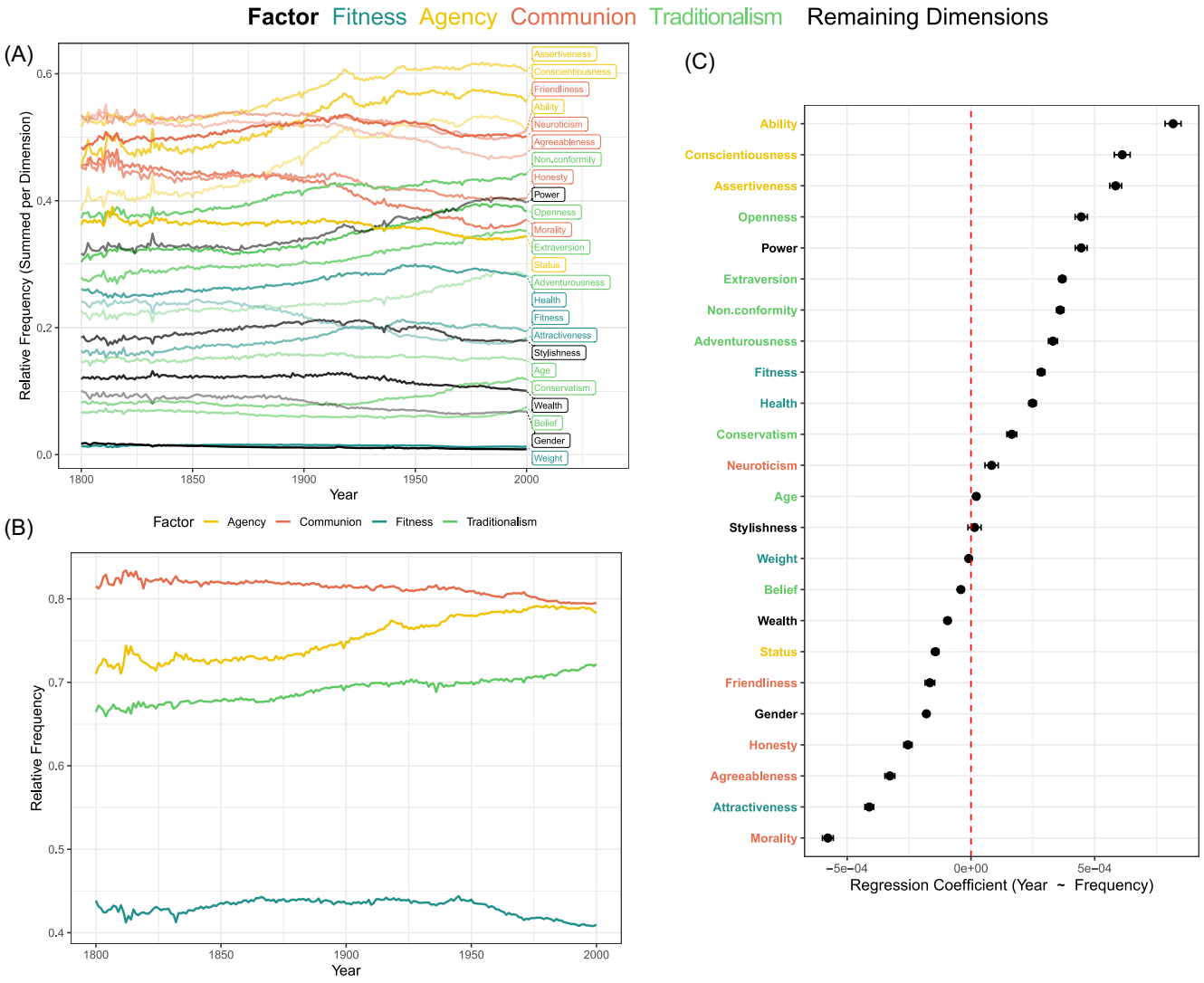
Changes in Trait Structure and Coherence Over History

Our final analyses focused on change in trait structure across the 19th and 20th centuries. There are many ways that trait space could

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 12

Trajectories of Frequency of Use of Each Dimension



Note. (A) The trend of frequency of use for each dimension from 1800 to 2000. The frequencies are relative to the total frequency of all trait vocabulary in a given year. (B) The trend of relative frequency for each factor from 1800 to 2000. (C) Ordered unstandardized regression coefficients using year to predict relative frequency. Error bars are 95% confidence intervals; red dashed line is 0. See the online article for the color version of this figure.

be changing, but we were specifically interested in coherence. Theoretically speaking, we focused on coherence because we sought to understand whether modern-day factors such as communion and agency have become more diffuse or concentrated over time. Would people in 1850 recognize these factors as coherent trait categories the way that we recognize them today?

Empirically speaking, we focused on coherence because we did not have human-subject ratings throughout history. The ratings used in Studies 2 and 3 came from a contemporary survey, which cannot be directly used in historical analyses.[4] Nonetheless, we can still assess coherence of a predefined list of words through geometrical features of historical word embeddings without human ratings.

Specifically, our method here relied on HistWords, which is a data set for analyzing language change using word vector embeddings (Hamilton et al., 2018). HistWords provides decadal embeddings of English vocabulary from the 1800s to the 1990s. We calculated centroids for each trait dimension (Charlesworth et al., 2022; Jackson et al., 2019; Kozlowski et al., 2019) and projected them into a common two-dimensional space using principal component analysis (Hotelling, 1933), applied jointly to the embeddings from all decades.

We then computed the Euclidean distance between each pair of dimensions within the same factor, averaged these distances at the factor level, and normalized them by decade. This within-factor distance served as a proxy for coherence: Lower values indicate that the dimensions within a factor are more semantically similar and potentially used more interchangeably. Utilizing this approach, we

[4] We still reported some historical analyses using these contemporary ratings in our Supplemental Material 10, which offer complementary insights into the historical change of valence and generality across trait factors.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

were able to analyze the historical change in the semantic coherence of the FACT factors: Did we see evidence that communion, agency, fitness, or traditionalism became more or less coherent over the last 200 years?

As shown in Figure 13, results[5] from this analysis suggested that communion dimensions have become more coherent over time, with a significant decline in average distance, $b = -3.34 \times 10^{-3}$, $SE = 5.48 \times 10^{-4}$, $p < .001$, $\beta = -0.82$, 95% CI [-4.49×10^{-3}, -2.19×10^{-3}]. By contrast, agency dimensions rose in their distance over time, $b = 2.69 \times 10^{-3}$, $SE = 4.32 \times 10^{-4}$, $p < .001$, $\beta = 0.83$, 95% CI [1.78×10^{-3}, 3.59×10^{-3}]. The other factors showed no significant change: fitness, $b = 0.79 \times 10^{-3}$, $SE = 8.33 \times 10^{-4}$, $p = .358$, $\beta = 0.22$, 95% CI [-0.96×10^{-3}, 2.54×10^{-3}], and traditionalism, $b = -0.13 \times 10^{-3}$, $SE = 6.77 \times 10^{-4}$, $p = .850$, $\beta = -0.05$, 95% CI [-1.55×10^{-3}, 1.29×10^{-3}].

We interpret these absolute slopes with caution because they may reflect changes in the Google Books Corpus. For example, there might be a global expansion or contraction of the Google Books Corpus, which is associated with its size or content diversity. Nevertheless, the relative differences between these slopes are meaningful because they control for corpus-level changes. Direct model comparisons using interaction terms revealed that communion showed a significantly more negative slope compared with agency ($b = -0.36$, $SE = 0.05$, $p < .001$, $\beta = -0.20$, 95% CI [−0.46, −0.25]), fitness ($b = -0.24$, $SE = 0.05$, $p < .001$, $\beta = -0.14$, 95% CI [−0.35, −0.14]), and traditionalism ($b = -0.19$, $SE = 0.05$, $p < .001$, $\beta = -0.11$, 95% CI [−0.30, −0.08]). Similarly, agency showed a significantly more positive slope compared with traditionalism ($b = 0.17$, $SE = 0.05$, $p = .003$, $\beta = 0.09$, 95% CI [0.06, 0.27]) and fitness ($b = 0.11$, $SE = 0.05$, $p = .039$, $\beta = 0.06$, 95% CI [0.01, 0.22]).

In the above analyses, trait coherence used distances that were standardized within each decade. This approach was chosen to account for potential shifts in the overall structure of the semantic embedding space over time—for instance, changes in model precision, corpus size, or language structure that could affect the absolute scale of pairwise distances. In Supplemental Material 6, we conducted analyses using raw average pairwise distances within each factor as the dependent variable. The results above generally remained unchanged.

Discussion

Study 4 demonstrated stability but also change in trait space. First, we saw remarkable stability in the relative number of words in each trait dimension and the relative frequency of word usage across trait dimensions. Specifically, communion, agency, and traditionalism all had larger and more frequently used vocabularies than fitness. Communion had the largest and most frequently used vocabulary, reflecting the primacy of communion dimensions like friendliness and honesty in social evaluation.

That being said, there was intriguing variation in the origination dates and frequency of use of these trait dimensions. The relative sequence of their emergence offers possible insights into their distinct evolutionary roles. For example, gender was the oldest dimension, resonating with its primacy in defining a human (Martin et al., 2024). In fact, all dimensions from the fitness factor tended to be relatively older. This may be because individual differences from these trait dimensions are more visible and objective, thus taking a more primal role in person perception, which is less tied to social changes (Gangestad & Scheyd, 2005). Communion dimensions—honesty and morality in particular—were also older, supporting their precedence in social evaluation (Eisenbruch & Krasnow, 2022).

By contrast, words about beliefs were comparatively young. Notably, words about conservatism were youngest—with an average "date of birth" around 1544—and they grew the fastest in the 14th century. This aligns with the relatively late formalization of the left–right political spectrum (Jost et al., 2009). Furthermore, dimensions such as openness, nonconformity, and extraversion were also relatively young. These traits connote individualism and cultural looseness (Stackhouse et al., 2024; Triandis, 2001), and their more recent emergence may reflect the rise of individualism and cultural looseness over the last 2 centuries (Jackson et al., 2019; Santos et al., 2017; Schulz et al., 2019), which increased variation in traits like openness and nonconformity. As people varied more along these dimensions, it became more necessary and meaningful to talk about them in everyday language.

The different trends in frequency across trait factors further suggest that people's use of trait language is linked to changes in social ecology. For example, the rise of agency language between the 1850s and 1920s occurred during the Second Industrial Revolution, in which drastic advances in technology created opportunities for skilled work in new domains. Similarly, the rise of traditionalism words in the latter half of the 20th century coincided with rising political polarization and partisanship in the English-speaking world (Iyengar & Westwood, 2015). It is possible that these trends in technology and political polarization have made capability and ideology more prominent features of person perception. On the other hand, words about communion are declining in frequency. Given the coarse nature of these data, it is difficult to make strong claims about why these trends coincide. The most provocative possibility is that these trends reflect "displacement" (e.g., that information about traditionalist aspects of character like political ideology is displacing information about warmth or attractiveness when people select partners). Future studies will need to test this possibility directly.

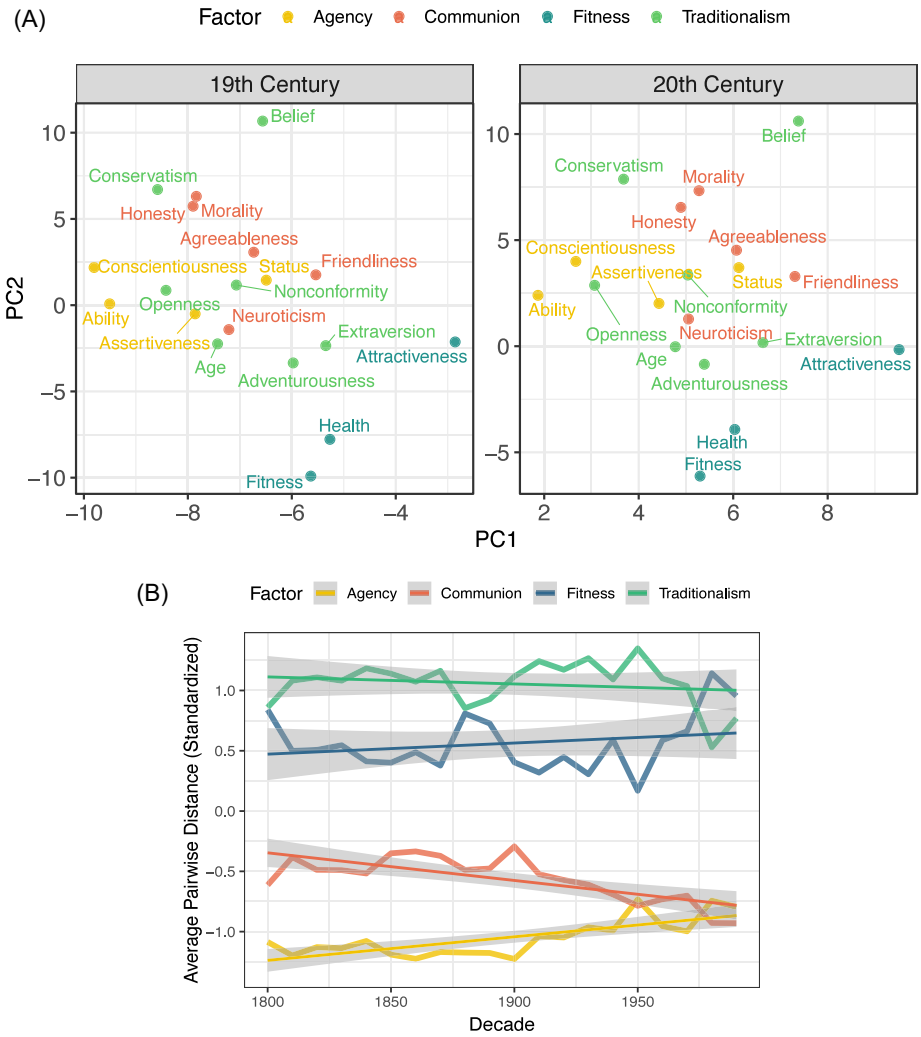
Finally, we found changes in trait structure that offer further evidence that trait language is linked to changes in social ecology. Our preliminary analyses suggested that trait structure has varied over time. Two notable changes include the rising coherence of communion traits and the falling coherence of agency traits. These findings echo two recent theories of history and trait categories. First, the diversification of agency trait space is consistent with the niche diversity hypothesis—that more occupational niches have corresponded to more trait categories (Durkee et al., 2022; Smaldino et al., 2019). Second, the rising coherence of communion traits is consistent with claims that communion dimensions, such as morality, are becoming more generalized as social networks are growing larger and more anonymous (Crockett et al., 2024; Jackson, Halberstadt, et al., 2023). At the same time, it is noteworthy that we rely here on the HistWords eng-all embeddings, which capture broad trends in the historical published record rather than a genre-balanced estimate of everyday interpersonal language use or underlying social evaluation. Accordingly, the historical links discussed here should be understood as suggestive interpretations of published language trends, not as direct evidence of the social–ecological processes that produced

[5] The results here are based on 23 dimensions with weight removed because it is markedly distant from the other dimensions, which would have disproportionately influenced the average pairwise-distance estimates. However, including this dimension did not change the results of the following semantic distance analyses. See Supplemental Material 5 for details.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 13

Historical Changes in Semantic Coherence of Each Trait Factor



Note. (A) The distributions of average semantic centroids of trait dimensions across the 19th and 20th centuries. (B) The trends of semantic coherence for each trait factor as depicted by average pairwise distance. The distance was standardized within each decade at the factor level. Shading represents 95% confidence intervals. PC1 = principal component 1; PC2 = principal component 2. See the online article for the color version of this figure.

them. Future research should continue to test these interpretations more directly.

General Discussion

Trait words offer a window into human psychology, serving as key tools for representing and communicating about individual differences. Across various psychological traditions, researchers have made major advances using lexicon-based approaches. However, much of this prior work faces persistent challenges—particularly in the incomplete sampling strategies and in their assumptions about the nature of human traits. These limitations have constrained the integration of different models and their broader applicability across contexts. In the present research, we sought to address these gaps through four studies.

In Study 1, we developed a large list of 2,847 English trait words using fine-tuned LLMs and GPT-based refinement. This list served as the empirical foundation for all subsequent analyses.

In Study 2, a total of 3,070 participants rated these words on 24 trait dimensions drawn from personality and social evaluation literatures. These ratings revealed a high-dimensional semantic structure that covered nearly all trait words and a lower dimensional four-factor FACT model. This FACT model exhibited satisfactory coverage of trait words. We also found that these factors varied in semantic coherence: Communion dimensions were most internally consistent and least entangled with other factors, whereas traditionalism and fitness traits were more dispersed and cross-loading.

In Study 3, we analyzed the evaluative structure of trait language. We found that valence closely tracked semantic direction for communion, agency, and fitness (except for weight) but was more

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

ambiguous for traditionalism. We also found that, although negative words tend to be more specific, the trait lexicon is not uniformly dominated by negative words. Rather, communion dimensions contained more negative words, whereas agency and traditionalism dimensions featured more positive words.

In Study 4, we traced the historical evolution of trait language. Although the relative size and frequency of trait dimensions remained largely stable, we observed systematic temporal shifts across the 19th and 20th centuries: Traditionalism words became more frequently used, while communion became less frequently used. Moreover, compared with other factors, communion dimensions became more semantically coherent, while agency dimensions became less coherent.

In what follows, we synthesize our findings across these studies, organized into four sections that highlight the functional constructionist nature of human trait language. These sections represent a preliminary effort to integrate our insights into a coherent theoretical model that acknowledges other theories while generating unique predictions.

Structure of Trait Space: High-Dimensional Nature and Functional Clusters

Traditional models of personality and social evaluation, such as the Big Five and the Big Two, have provided influential and useful accounts of how people organize information about individual differences (Abele et al., 2021; McCrae & Costa, 1997). These models were developed to answer particular theoretical questions about personality and impression formation. Building on these traditions, our goal was to ask what becomes visible when the object of analysis is broadened to the wider space of stable, comparable, and socially meaningful person descriptors. From this broader perspective, recent work suggests that low-dimensional structures in person perception can be understood as socially functional summaries of a richer, high-dimensional representational space (Freeman & Lin, 2025). The organization of this space may vary across the kinds of targets being judged (Collova et al., 2019; Jones et al., 2021; C. Lin & Thornton, 2023). It may also vary across cultures and historical periods, as suggested by cross-cultural trait research and diachronic semantic analyses (Aghababaei et al., 2025; Cheung et al., 2001; Gurven et al., 2013; Jackson, Halberstadt, et al., 2023; Thalmayer et al., 2020, 2021).

Consistent with this view, our findings suggest that trait space is inherently high dimensional. Using our representative and extensive English trait lexicon, we found that the 24 dimensions we synthesized from prior literature accounted for the vast majority of trait words across a wide range of relevance thresholds. Even after reducing overlap between dimensions, a comparatively large number of dimensions was still needed to adequately cover most of the lexicon. In Supplemental Material 4, we further demonstrate the unique contribution of these dimensions using a principled stepwise dropout procedure. Together, these results suggest that our trait perception system is structured around a richly organized, high-dimensional semantic network.

Although we also derived a more parsimonious component structure—the FACT model—through factor reduction techniques, we do not interpret these components as direct biological mappings. We also do not assume that the FACT factors are fixed or universal. Rather, we view these components as loosely organized functional clusters, each supporting a core psychological function that recurs in everyday social life. The specific trait dimensions within each cluster may vary across cultures and evolve over time. For example, while Traditionalism represents a shared theme in social evaluation, its manifestations may vary across cultures. In English-speaking societies, openness serves as a salient expression of traditionalism, but past work shows that it is not prominent as an independent dimension in Chinese or Iranian populations. Instead, culturally specific traits like Renqing (roughly, relationship orientation) and Azadmanesh (roughly, high-mindedness) play similar functional roles in Chinese and Persian contexts, respectively (Aghababaei et al., 2025; Cheung et al., 2001). We believe that this functional perspective offers a more integrative and generative framework for interpreting both our findings and prior lexicon-based trait research—an idea we elaborate in the sections that follow.

Differentiating Communion and Traditionalism: A Game-Theoretic Perspective

Among the agency, communion, and traditionalism factors in our model, the agency and communion factors closely replicate those found in classic two-factor models (Abele et al., 2021). The key distinction lies in how our model distinguishes communion dimensions (e.g., warmth, morality) from personal preference dimensions (e.g., beliefs, openness), which are often grouped together in prior work. For example, Thalmayer et al. (2025) recently proposed a cross-cultural Big Two model that clustered warmth and morality with traits like obedient/disobedient and peaceful/hot-tempered under the broad category of social self-regulation. By contrast, our empirical analyses revealed that communion and traditionalism dimensions are distinct: They emerged separately in our EFA and demonstrated systematically different properties.

Specifically, our analyses revealed that communion was the most semantically coherent factor, while traditionalism was the least coherent. We also found that people agreed strongly about the positivity of communion traits but did not agree about the positivity of traditionalism traits. Moreover, although communion dimensions were more frequently used overall, their usage has been declining over time, whereas several traditionalism dimensions have become increasingly frequent over the past two centuries.

In our previous discussion in Study 3, we interpreted these differences through a need-based framework, suggesting that communion reflects universally shared basic needs, while traditionalism represents more context-dependent, secondary needs that vary across communities or individuals. While useful, this approach rests on assumptions about which needs are universal, and these assumptions remain theoretically contested and difficult to disprove. To avoid strong assumptions about basic needs, we offer a complementary perspective grounded in game theory (Colman, 2003).

From a game-theoretic perspective, person perception often depends on at least three broad kinds of information: whether others are likely to help or harm us, whether their preferences align with ours, and whether they are capable of effectively carrying out their intentions. These concerns map roughly onto cooperation (Axelrod & Hamilton, 1981), coordination (Cooper et al., 1992), and potency (Osgood, 1964), respectively. Although all three may matter in social life, they need not be equally central to all parts of trait space.

Within this lens, communion dimensions may be especially informative for cooperation concerns, whereas traditionalism

dimensions may be more informative for coordination concerns. This contrast provides one possible way to understand why these dimensions, although sometimes grouped together in prior frameworks, appear more sharply differentiated in our data. In particular, cooperation-relevant traits may invite broader semantic consensus and more stable evaluative agreement, whereas coordination-relevant traits may remain more heterogeneous because their value depends more strongly on local norms, goals, and alignments.

We present this game-theoretic perspective as a speculative and interpretive lens while emphasizing that the studies in this article do not directly test this perspective. Rather, its value lies in organizing the observed patterns and generating more precise hypotheses for future work. For example, future studies could test whether communion traits are used more in cooperation-relevant judgments, whereas traditionalism traits are used more in coordination-relevant judgments or whether these dimensions differ systematically in how much evaluative consensus they elicit across social contexts.

Perceptual Foundations of Trait Language: The Unique Role of Fitness

In addition to the agency, communion, and traditionalism factors, our analyses revealed a distinct fitness factor, comprising dimensions such as attractiveness, fitness, health, and weight. These traits tend to reflect relatively objective and physically observable characteristics—making them among the most immediately accessible cues in social perception. Although theory-driven models have often ignored fitness-related traits intentionally because they do not meet the prerequisite of being individual differences in internal psychology or behavior, these traits displayed several patterns in our data that indicated their unique functions in social evaluation.

First, historical analyses showed that fitness traits tended to emerge earlier than other dimensions, suggesting their foundational role in trait representation. These traits exhibited low within-factor generality, likely because their accessibility reduces the need for inferential links among them (e.g., it is not necessary to infer weight from health). By contrast, their between-factor generality was among the highest, perhaps indicating that their perceptual accessibility makes them a common basis for inferring more abstract psychological traits. This inference tendency was amplified by positive valence, which selectively increased between-factor generality, but not within-factor coherence. Notably, while traits like attractive, fit, and healthy consistently carried positive valence, people showed little agreement on the valence of overweight—echoing shifting cultural attitudes toward body weight in English-speaking contexts from positive to negative over time (Nutter et al., 2024).

These findings align with the view that visual physical cues serve as proximal inputs to social evaluation, forming a perceptual foundation upon which more abstract, inferential judgments are built (Freeman & Lin, 2025). In this sense, fitness traits are especially likely to serve as heuristics (Gigerenzer & Gaissmaier, 2011): fast, frugal, often oversimplified, yet functionally useful in guiding social judgments (C. Lin et al., 2025). Sometimes, these heuristics may reflect actual evolved group-level correlations. For example, greater prosociality may have coevolved with more feminized or juvenilized facial features (Hare, 2017). Other times, such inferences may reflect universal or culturally specific overgeneralizations (Sutherland et al., 2018; Zebrowitz, 2017). For instance, avoidance of "unfit" faces may reflect an evolved bias to avoid disease or genetic anomalies (Zebrowitz et al., 2003), while in the U.S. context, tall Black men may be overgeneralized as threatening due to histories of discrimination (Hester & Gray, 2018).

Importantly, fitness traits may not always play a strong role in social evaluation. Although often fast and automatic, reliance on appearance cues in impression formation can sometimes be deliberately suppressed through interventions and training (Hong et al., 2024; Roy et al., 2025) or normatively downweighted in contexts with strong egalitarian norms that explicitly sanction appearance-based judgments (Martin & North, 2022). Thus, we view fitness as playing a primal role in social evaluation—rooted in early evolutionary history and perceptual heuristics, yet malleable, open to cultural construction and personal or societal regulation.

Taken together, we argue that fitness should be treated as a separate factor in trait representation. Rather than serving merely as a record of internal dispositional traits, it represents a basic layer of social perception—one that grounds and scaffolds the more abstract inferences we make about others. Nonetheless, we recognize that some readers may be primarily interested in the more narrowly dispositional portion of trait space. For such purposes, our data infrastructure can also readily support analyses focused only on the agency, communion, and traditionalism dimensions.

Coevolution of Trait Language With Social Ecology

Our study not only documents historical changes in trait concepts but also offers a framework for interpreting how such changes may relate to shifts in social ecology. From a functional constructionist perspective, we view trait language as a shared representational system for encoding meaningful individual differences that supports person perception in everyday social life, enabling us to make "good-enough" decisions to remain adaptive in changing social ecology. Therefore, the representational space of trait language should evolve in tandem with the changing demands of the social environment.

In our study, this evolution manifests in several ways: increases or decreases in the number and frequency of trait words, as well as expansions or contractions within existing semantic dimensions. Future research may even provide evidence for the emergence of new dimensions over time. Importantly, just as traits are not direct reflections of biological modules as assumed by universalist perspectives, they are also not mere mirrors of the observable landscape of individual differences. Rather, we reason that trait concepts are strategically constructed, jointly shaped by a dynamic interplay of biological constraints, population-level variation, and informational ecology, to maintain their functionality in social perception.

One key factor shaping trait semantics is the objective landscape of individual differences, which provides the raw material for social learning. As this landscape changes, so too does the conceptual system we use to represent it (Piaget, 1952; Tenenbaum et al., 2011). Our findings are consistent with this dynamic-alignment account by documenting systematic shifts in the trait lexicon over time. For example, we observed that the 16th century featured the largest increase in trait word emergence across dimensions, coinciding with historical developments such as religious reform, early nation-state formation, and cultural flourishing in early modern England (James, 1988). These events likely produced a broader spectrum of social roles and interpersonal expectations, requiring a more nuanced

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

conceptual vocabulary to describe individual variation. This pattern aligns with the niche diversity hypothesis, which posits that the enrichment of social ecological niches through processes such as occupational diversification and increased tolerance toward minorities and dissenters leads to greater variation in human traits and, correspondingly, a more differentiated conceptual structure (Durkee et al., 2022; Montiglio et al., 2013; Smaldino et al., 2019).

However, the structure of trait semantics is not determined solely by the objective landscape of individual differences but also by how people are able to perceive, store, and use trait information within their informational ecology. As societies expand and social networks grow, individuals come into contact with a larger number of people but have less time and fewer cognitive resources to devote to each one (Dunbar, 1993). The rising anonymity of social networks may increase the functional value—and therefore the use—of trait words as information-light inference heuristics. For example, we may need to infer whether Target X has Trait B from our observation of Trait A when Trait B is relevant to a decision but Trait A is all we know. Over time, such inferential habits may cause their meanings to converge semantically.

This pressure toward semantic generalization is selective. It depends on whether the traits in question share a stable underlying relationship. Traits associated with cooperation, such as warmth, honesty, and generosity, are more likely to be generalized because they stem from a common prosocial orientation and tend to covary across contexts (Eisenbruch & Krasnow, 2022; Hare, 2017). By contrast, traits associated with agency, such as special skills, are more context specific and often orthogonal. Generalizing across them may introduce errors that reduce the utility of such heuristics. Our findings are consistent with this asymmetry: Communion traits showed greater semantic convergence over time, while agency traits retained more distinct boundaries. These patterns suggest that the evolution of trait semantics reflects not only what people need to track but also what they are able to infer reliably under real-world cognitive and informational constraints.

This aligns with recent work on social cognitive maps (FeldmanHall et al., 2025), which suggests that humans abstract over relational information to navigate dense and dynamic social networks more effectively. Such abstraction does not reflect veridical encoding but rather strategically compressed representations that prioritize inference utility over precision—enabling individuals to predict social links and information flow with limited cognitive resources.

Limitations and Future Directions

While these findings are promising, much work remains to further refine and build upon these initial steps. Here, we underscore four key limitations and future directions.

First, our analysis is restricted to English-speaking contexts, and it remains unclear how well our findings generalize across languages and societies. We encourage future studies to apply our pipeline to the investigation of other languages. Our approach provides an easily portable paradigm: directly investigating the full range of trait words and native trait dimensions in a target language and tracking their evolution over time. This may help avoid the inappropriate imposition of English-centric trait structures on other languages and enable researchers to test whether functionally similar factors—such as those captured in our FACT model—emerge across cultures and whether their generality, valence, and temporal dynamics show comparable patterns. Such cross-linguistic testing is critical for evaluating the cultural generalizability of our framework. Moreover, even within a single language such as English, general-purpose LLMs are likely to disproportionately reflect evaluative norms from heavily represented contexts (e.g., the United States or the United Kingdom), potentially underweighting trait representations from other English-speaking populations. Future work should therefore combine our pipeline with more culturally sensitive, within-language approaches—for example, models trained on country-specific corpora (Souza et al., 2020) or prompts that explicitly anchor judgments in a given national context (Lu et al., 2025).

Second, our analyses are limited in several time-related ways. First, the corpora we use place hard bounds (the 19th and 20th centuries) on the temporal window for studying semantic change, although we supplement these with frequency trends based on annual counts that extend to 2022. As diachronic embedding methods and corpora continue to improve, future work could extend our semantic analyses beyond the 20th century, which could allow tests of whether trends in communion and traditionalism language reverse in the post-2010 era of social media, ideological polarization, and LLM-mediated communication. Second, our analysis relies on a trait lexicon compiled from modern English sources. Strictly speaking, this limits our ability to reconstruct the true historical "trait space" given potential semantic drift and the emergence or obsolescence of trait concepts over time (Hamilton et al., 2018). To partially mitigate this concern, we sampled candidate trait words from across the 1800–2008 period in the Google Books corpus, ensuring that the lexicon includes historically prevalent words. Future work could build on this using more temporally sensitive methods—for example, training semantic models on historical corpora to capture period-specific meanings in trait classification or applying a contextualized construct representation approach to identify our trait list's culturally appropriate equivalents in older texts (Chen et al., 2024). Such tools may allow researchers to reconstruct historical trait vocabularies more accurately and test how semantic structure and dimensionality have shifted across time periods. Finally, we note that no text-based approach can fully guarantee representativeness of everyday interpersonal language use, especially in earlier periods when literacy was lower (Cressy, 1977) and surviving written materials disproportionately reflect the language of more educated writers (Varnum et al., 2024). Future work could also triangulate our text-based trends with complementary nontext cultural artifacts (e.g., portraits and other visual records) to strengthen historical inference (Baumard et al., 2024).

Third, due to the exploratory nature of our study, many of our analyses are descriptive or correlational. While sufficient for answering our initial questions, more rigorous experimental or simulation-based approaches are necessary to establish causal relationships. For example, we put forth a game-theoretic perspective to understand the components of human trait space, which provides a formalizable structure for such efforts. Future studies could use agent-based modeling or lab-based designs to test specific predictions derived from our framework. For example, can greater consensus on communion traits relative to traditionalism traits be replicated in a repeated-game setting as a direct function of differing payoff matrices? Are communion traits more likely to be used to infer other traits within the same domain, while traditionalism traits are more cross-dimensional? Does expansion of a simulated social network increase the use of communion-based inferences more than others? Controlled tests of these propositions will help verify the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

causal mechanisms of our theoretical model. A complementary direction for causal tests would be to identify the more specific psychologically active ingredients in social ecologies that underlie historical changes in trait words. For example, future work could examine whether shifts in social coordination demands, partner-selection pressures, institutional complexity, or information environments play distinct roles in reshaping trait semantics over time.

Last, our analyses relied on aggregated ratings from a large, convenience sample of native English speakers, which inevitably smooths over potentially meaningful variation across social groups. For example, we did not directly measure participants' political orientation as a study-specific variable. Because Prolific samples tend to be liberal leaning, our aggregate ratings—particularly for traditionalism-related words—may not fully represent the ideological diversity of English-speaking populations. This limitation is important because trait representations are not monolithic (Hehman et al., 2017). As recent work suggests, on the one hand, different populations may construct and interpret trait words in systematically different ways depending on their ideological orientation or lived experience (Freeman & Lin, 2025; Kura & Rentzsch, 2025). On the other hand, the same person's representation of trait space toward different targets may also vary with their unique interaction dynamics (Yzerbyt et al., 2025). For instance, compared with political moderates, extremists may associate communion dimensions more with traditionalism dimensions (Iyengar et al., 2019). Furthermore, advocates of egalitarianism may be less likely to evaluate people on the basis of their fitness dimensions (e.g., weight) but likely to show some forms of ageism (Martin & North, 2022). Similarly, racial and ethnic minority groups may differ in the generality or valence they assign to fitness traits (e.g., overweight, skinny) shaped by historical stigma or identity norms (Hester & Hehman, 2023). These group-specific trait appraisals are not just a source of noise but may reflect deeply rooted differences in motivational priorities, social affordances, or historical narratives (Kay et al., 2023; Neuberg et al., 2020). Importantly, our framework—based on a large vocabulary and a dense, multidimensional annotation procedure—is particularly well suited for such comparisons. Future work could further extend our approach to examine group-specific trait-representation networks, which may help uncover how identity, ideology, and experience shape fundamental social evaluations.

In addition to capturing how different human populations interpret trait words, future research may further explore differences between how humans and machines interpret trait words. To this end, we present an exploratory analysis in Supplemental Material 8, in which we replicated key analyses using dimensional ratings simulated by GPT rather than human raters. While we generally reproduce our findings regarding valence, trait age, and frequency, we also found meaningful discrepancies in the generality and semantic coherence, which resonate well with recent investigations on the response style of LLMs (Nadler et al., 2023; A. Wang et al., 2025). We view these GPT-based analyses as a proof of concept that our pipeline can be applied to different "observers," such as different human groups or different AI models.

Conclusion

Here we develop one of the most comprehensive English trait lexicons to date using LLMs and human annotation and use this lexicon to explore a functional constructionist account of trait language. In this account, trait words are adaptive and flexible tools for navigating our social environments. The structure of trait words provides a basis for social inference across contexts and behaviors, the valence of trait words provides a basis for social comparison and evaluation, and historical changes in trait words allow trait space to adaptively change alongside social ecology.

Our studies show that English trait language is organized in a high-dimensional space that can be partially reduced to four functionally distinct factors—fitness, agency, communion, and traditionalism. These factors differ in their possible social functions and have unique distributions of semantic coherence and valence. Historical analyses further reveal stability and change in trait concepts over time, as social ecologies and informational demands evolve. Together, these findings suggest that trait words do more than describe people—they are active products of the social environments and interactions that they serve to structure and predict.

References

Abdurahman, S., Atari, M., Karimi-Malekabadi, F., Xue, M. J., Trager, J., Park, P. S., Golazizian, P., Omrani, A., & Dehghani, M. (2024). Perils and opportunities in using large language models in psychological research. *PNAS Nexus*, *3*(7), Article pgae245. https://doi.org/10.1093/pnasnexus/pgae245

Abdurahman, S., Ziabari, A. S., Moore, A., Bartels, D., & Dehghani, M. (2024). *Evaluating large language models in psychological research: A guide for reviewers*. OSF. https://doi.org/10.31234/osf.io/ag7hy

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314. https://doi.org/10.1037/rev0000262

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in Psychology*, *7*, Article 1810. https://doi.org/10.3389/fpsyg.2016.01810

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*(5), 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Aghababaei, N., Saucier, G., Vader, V., & Razavi, P. (2025). Comprehensive personality structure in the Persian language: High-dimensionality analyses of trait adjectives. *Journal of Personality and Social Psychology*, *129*(1), 135–151. https://doi.org/10.1037/pspp0000557

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), i–171. https://doi.org/10.1037/h0093360

Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences*, *21*(2), 69–79. https://doi.org/10.1016/j.tics.2016.12.006

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258–290. https://doi.org/10.1037/h0055756

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*(2), 139–152. https://doi.org/10.1177/1088868314523838

Averill, J. R. (1980). On the paucity of positive emotions. In K. R. Blankstein, P. Pliner, & J. Polivy (Eds.), *Assessment and modification of emotional behavior* (pp. 7–45). Springer. https://doi.org/10.1007/978-1-4684-3782-9_2

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Bao, H.-W.-S. (2024). The Fill-Mask Association Test (FMAT): Measuring propositions in natural language. *Journal of Personality and Social Psychology*, *127*(3), 537–561. https://doi.org/10.1037/pspa0000396

Barabucci, G., Shia, V., Chu, E., Harack, B., Laskowski, K., & Fu, N. (2024). Combining multiple large language models improves diagnostic accuracy. *NEJM AI*, *1*(11), Article AIcs2400502. https://doi.org/10.1056/AIcs2400502

Baumard, N., Safra, L., Martins, M., & Chevallier, C. (2024). Cognitive fossils: Using cultural artifacts to reconstruct psychological changes throughout history. *Trends in Cognitive Sciences*, *28*(2), 172–186. https://doi.org/10.1016/j.tics.2023.10.001

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Benet, V., & Waller, N. G. (1995). The Big Seven factor model of personality description: Evidence for its cross-cultural generality in a Spanish sample. *Journal of Personality and Social Psychology*, *69*(4), 701–718. https://doi.org/10.1037/0022-3514.69.4.701

Benet-Martínez, V., & Waller, N. G. (1997). Further evidence for the cross-cultural generality of the Big Seven factor model: Indigenous and imported Spanish personality constructs. *Journal of Personality*, *65*(3), 567–598. https://doi.org/10.1111/j.1467-6494.1997.tb00327.x

Bjornsdottir, R. T., & Rule, N. O. (2017). The visibility of social class from facial cues. *Journal of Personality and Social Psychology*, *113*(4), 530–546. https://doi.org/10.1037/pspa0000091

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, *17*(5), 475–482. https://doi.org/10.1177/0146167291175001

Brewer, M. B. (2007). The importance of being we: Human nature and intergroup relations. *American Psychologist*, *62*(8), 728–738. https://doi.org/10.1037/0003-066X.62.8.728

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology*, *38*(4), 476–506. https://doi.org/10.1037/h0054116

Charlesworth, T. E. S., Caliskan, A., & Banaji, M. R. (2022). Historical representations of social groups across 200 years of word embeddings from Google Books. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(28), Article e2121798119. https://doi.org/10.1073/pnas.2121798119

Chen, Y., Li, S., Li, Y., & Atari, M. (2024). *Surveying the dead minds: Historical-psychological text analysis with contextualized construct representation (CCR) for classical Chinese*. arXiv. https://arxiv.org/abs/2403.00509

Cheung, F. M., Leung, K., Zhang, J.-X., Sun, H.-F., Gan, Y.-Q., Song, W.-Z., & Xie, D. (2001). Indigenous Chinese personality constructs: Is the five-factor model complete? *Journal of Cross-Cultural Psychology*, *32*(4), 407–433. https://doi.org/10.1177/0022022101032004003

Church, A. T. (2000). Culture and personality: Toward an integrated cultural trait psychology. *Journal of Personality*, *68*(4), 651–703. https://doi.org/10.1111/1467-6494.00112

Collova, J. R., Sutherland, C. A. M., & Rhodes, G. (2019). Testing the functional basis of first impressions: Dimensions for children's faces are not the same as for adults' faces. *Journal of Personality and Social Psychology*, *117*(5), 900–924. https://doi.org/10.1037/pspa0000167

Colman, A. M. (2003). Beyond rationality: Rigor without mortis in game theory. *Behavioral and Brain Sciences*, *26*(2), 180–198. https://doi.org/10.1017/S0140525X0350005X

Connor, P., Nicolas, G., Antonoplis, S., & Koch, A. (2026). Unconstrained descriptions of Facebook profile pictures support high-dimensional models of impression formation. *Personality and Social Psychology Bulletin*, *52*(1), 53–69. https://doi.org/10.1177/01461672241266651

Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. W. (1992). Communication in coordination games. *The Quarterly Journal of Economics*, *107*(2), 739–771. https://doi.org/10.2307/2118488

Cressy, D. (1977). Literacy in seventeenth-century England: More evidence. *Journal of Interdisciplinary History*, *8*(1), 141–150. https://doi.org/10.2307/202599

Crockett, M. J., Kim, J. S., & Shin, Y. S. (2024). Intuitive theories and the cultural evolution of morality. *Current Directions in Psychological Science*, *33*(4), 211–219. https://doi.org/10.1177/09637214241245412

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., … Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), 1–33. https://doi.org/10.1348/014466608X314935

Cutler, A., & Condon, D. M. (2023). Deep lexical hypothesis: Identifying personality structure in natural language. *Journal of Personality and Social Psychology*, *125*(1), 173–197. https://doi.org/10.1037/pspp0000443

Dawson, S. K., Carmona, C. P., González-Suárez, M., Jönsson, M., Chichorro, F., Mallen-Cooper, M., Melero, Y., Moor, H., Simaika, J. P., & Duthie, A. B. (2021). The traits of "trait ecologists": An analysis of the use of trait and functional trait terminology. *Ecology and Evolution*, *11*(23), 16434–16445. https://doi.org/10.1002/ece3.8321

Delton, A. W., & Robertson, T. E. (2012). The social cognition of social foraging: Partner selection by underlying valuation. *Evolution and Human Behavior*, *33*(6), 715–725. https://doi.org/10.1016/j.evolhumbehav.2012.05.007

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv. https://arxiv.org/abs/1810.04805

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*(6), 1138–1151. https://doi.org/10.1037/0022-3514.91.6.1138

DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the Big Five. *Psychological Science*, *21*(6), 820–828. https://doi.org/10.1177/0956797610370159

Dias, N., & Lelkes, Y. (2022). The nature of affective polarization: Disentangling policy disagreement from partisan identity. *American Journal of Political Science*, *66*(3), 775–790. https://doi.org/10.1111/ajps.12628

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*(6), 1246–1256. https://doi.org/10.1037/0022-3514.73.6.1246

Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research*, *16*(2), 149–170. https://doi.org/10.1207/s15327906mbr1602_2

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, *18*(3), Article e0279720. https://doi.org/10.1371/journal.pone.0279720

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–694. https://doi.org/10.1017/S0140525X00032325

Dupree, C. H., & Fiske, S. T. (2017). Universal dimensions of social signals: Warmth and competence. In J. K. Burgoon, N. Magnenat-Thalmann, M. Pantic, & A. Vinciarelli (Eds.), *Social signal processing* (pp. 23–33). Cambridge University Press. https://doi.org/10.1017/9781316676202.003

Durkee, P. K., Lukaszewski, A. W., Von Rueden, C. R., Gurven, M. D., Buss, D. M., & Tucker-Drob, E. M. (2022). Niche diversity predicts personality structure across 115 nations. *Psychological Science*, *33*(2), 285–298. https://doi.org/10.1177/09567976211031571

Dweck, C. S. (2017). From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychological Review*, *124*(6), 689–719. https://doi.org/10.1037/rev0000082

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but … : A meta-analytic review of research on

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Eisenbruch, A. B., & Krasnow, M. M. (2022). Why warmth matters more than competence: A new evolutionary approach. *Perspectives on Psychological Science*, *17*(6), 1604–1623. https://doi.org/10.1177/17456916211071087

Ellemers, N., Pagliaro, S., & Barreto, M. (2013). Morality and behavioural regulation in groups: A social identity approach. *European Review of Social Psychology*, *24*(1), 160–193. https://doi.org/10.1080/10463283.2013.841490

Feher, A., & Vernon, P. A. (2021). Looking beyond the Big Five: A selective review of alternatives to the Big Five model of personality. *Personality and Individual Differences*, *169*, Article 110002. https://doi.org/10.1016/j.paid.2020.110002

FeldmanHall, O., Son, J.-Y., & Bhandari, A. (2025). Abstract cognitive maps for complex social systems. *Current Directions in Psychological Science*, *34*(6), 349–356. https://doi.org/10.1177/09637214251342742

Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, *63*(6), 877–889. https://doi.org/10.1037/0022-3514.63.6.877

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Freeman, J. B., & Lin, C. (2025). A high-dimensional model of social impressions. *Trends in Cognitive Sciences*, *29*(9), 790–801. https://doi.org/10.1016/j.tics.2025.04.011

Galton, F. (1884). Measurement of character. *Fortnightly Review*, *36*(212), 179–185.

Gangestad, S. W., & Scheyd, G. J. (2005). The evolution of human physical attractiveness. *Annual Review of Anthropology*, *34*(1), 523–548. https://doi.org/10.1146/annurev.anthro.33.070203.143733

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Gosling, S. D., & John, O. P. (1999). Personality dimensions in nonhuman animals: A cross-species review. *Current Directions in Psychological Science*, *8*(3), 69–75. https://doi.org/10.1111/1467-8721.00017

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*(1), 3–25. https://doi.org/10.1037/0033-295X.109.1.3

Gu, K., Tuecke, E., Katz, D., Horesh, R., Alvarez-Melis, D., & Yurochkin, M. (2024). *CharED: Character-wise ensemble decoding for large language models* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2407.11009

Gurven, M., Von Rueden, C., Massenkoff, M., Kaplan, H., & Lero Vie, M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian Amazon. *Journal of Personality and Social Psychology*, *104*(2), 354–370. https://doi.org/10.1037/a0030841

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2018). *Diachronic word embeddings reveal statistical laws of semantic change*. arXiv. https://arxiv.org/abs/1605.09096

Hare, B. (2017). Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual Review of Psychology*, *68*(1), 155–186. https://doi.org/10.1146/annurev-psych-010416-044201

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847–853. https://doi.org/10.1111/j.1467-9280.2006.01793.x

Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in mPFC. *Social Cognitive and Affective Neuroscience*, *2*(1), 45–51. https://doi.org/10.1093/scan/nsl037

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*(1), 81–91. https://doi.org/10.1037/0022-3514.78.1.81

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513–529. https://doi.org/10.1037/pspa0000090

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, *12*(3), 123–135. https://doi.org/10.1002/evan.10110

Hester, N., & Gray, K. (2018). For Black men, being tall increases threat stereotyping and police stops. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2711–2715. https://doi.org/10.1073/pnas.1714454115

Hester, N., & Hehman, E. (2023). Dress is a fundamental component of person perception. *Personality and Social Psychology Review*, *27*(4), 414–433. https://doi.org/10.1177/10888683231157961

Hofstede, G., Neuijen, B., Ohayv, D. D., & Sanders, G. (1990). Measuring organizational cultures: A qualitative and quantitative study across twenty cases. *Administrative Science Quarterly*, *35*(2), 286–316. https://doi.org/10.2307/2393392

Hong, Y., Chua, K.-W., & Freeman, J. B. (2024). Reducing facial stereotype bias in consequential social judgments: Intervention success with White male faces. *Psychological Science*, *35*(1), 21–33. https://doi.org/10.1177/09567976231215238

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441. https://doi.org/10.1037/h0071325

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, *59*(3), 690–707. https://doi.org/10.1111/ajps.12152

Jackson, J. C., Gelfand, M., De, S., & Fox, A. (2019). The loosening of American culture over 200 years is associated with a creativity–order trade-off. *Nature Human Behaviour*, *3*(3), 244–250. https://doi.org/10.1038/s41562-018-0516-z

Jackson, J. C., Halberstadt, J., Takezawa, M., Liew, K., Smith, K., Apicella, C., & Gray, K. (2023). Generalized morality culturally evolves as an adaptive heuristic in large social networks. *Journal of Personality and Social Psychology*, *125*(6), 1207–1238. https://doi.org/10.1037/pspa0000358

Jackson, J. C., Lindquist, K., Drabble, R., Atkinson, Q., & Watts, J. (2023). Valence-dependent mutation in lexical evolution. *Nature Human Behaviour*, *7*(2), 190–199. https://doi.org/10.1038/s41562-022-01483-8

James, M. (1988). *Society, politics and culture: Studies in early modern England*. Cambridge University Press. https://doi.org/10.1017/CBO9780511560613

Jiang, D., Ren, X., & Lin, B. Y. (2023). *LLM-blender: Ensembling large language models with pairwise ranking and generative fusion* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2306.02561

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution*, *28*(8), 474–481. https://doi.org/10.1016/j.tree.2013.05.014

Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxsom, N. G., Lewis, S. C., Foroni, F.,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Willis, M. L., Cubillas, C. P., Vadillo, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., … Coles, N. A. (2021). To which world regions does the valence–dominance model of social perception apply? *Nature Human Behaviour*, *5*(1), 159–169. https://doi.org/10.1038/s41562-020-01007-2

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual Review of Psychology*, *60*(1), 307–337. https://doi.org/10.1146/annurev.psych.60.110707.163600

Kawamura, Y., & Barclay, P. (2025). Wealth or generosity? People choose partners based on whichever is more variable. *Evolution and Human Behavior*, *46*(5), Article 106727. https://doi.org/10.1016/j.evolhumbehav.2025.106727

Kay, A. C., Ponce de Leon, R., Ho, A. K., & Kteily, N. S. (2023). Motivated egalitarianism. *Current Directions in Psychological Science*, *32*(4), 293–299. https://doi.org/10.1177/09637214231154810

Koch, A., Dorrough, A., Glöckner, A., & Imhoff, R. (2020). The ABC of society: Perceived similarity in agency/socioeconomic success and conservative–progressive beliefs increases intergroup cooperation. *Journal of Experimental Social Psychology*, *90*, Article 103996. https://doi.org/10.1016/j.jesp.2020.103996

Koch, A., Dotsch, R., Imhoff, R., Unkelbach, C., & Alves, H. (2025). Ideological beliefs as cues to exploitation-exploration behavior. *Journal of Experimental Social Psychology*, *117*, Article 104718. https://doi.org/10.1016/j.jesp.2024.104718

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675–709. https://doi.org/10.1037/pspa0000046

Koch, A., Kervyn, N., Kervyn, M., & Imhoff, R. (2018). Studying the cognitive map of the U.S. states: Ideology and prosperity stereotypes predict interstate prejudice. *Social Psychological and Personality Science*, *9*(5), 530–538. https://doi.org/10.1177/1948550617715070

Koch, A., Smith, A., Fiske, S. T., Abele, A. E., Ellemers, N., & Yzerbyt, V. (2024). Validating a brief measure of four facets of social evaluation. *Behavior Research Methods*, *56*(8), 8521–8539. https://doi.org/10.3758/s13428-024-02489-y

Koch, A., Yzerbyt, V., Abele, A., Ellemers, N., & Fiske, S. T. (2021). Social evaluation: Comparing models across interpersonal, intragroup, intergroup, several-group, and many-group contexts. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 63, pp. 1–68). Academic Press. https://doi.org/10.1016/bs.aesp.2020.11.001

Kozlowski, A. C., Kwon, H., & Evans, J. A. (2024). *In silico sociology: Forecasting COVID-19 polarization with large language models* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2407.11190

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, *84*(5), 905–949. https://doi.org/10.1177/0003122419877135

Kura, M., & Rentzsch, K. (2025). Context effects on the personality network—Exploring differences between East and West Germans. *Journal of Personality and Social Psychology*, *129*(4), 736–758. https://doi.org/10.1037/pspp0000569

Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, *80*(S1), 392–410. https://doi.org/10.1093/poq/nfw005

Leung, K., & Bond, M. H. (2004). Social axioms: A model for social beliefs in multicultural perspective. *Advances in Experimental Social Psychology*, *36*, 119–197. https://doi.org/10.1016/S0065-2601(04)36003-X

Lin, C., Keles, U., Thornton, M. A., & Adolphs, R. (2025). How trait impressions of faces shape subsequent mental state inferences. *Nature Human Behaviour*, *9*(1), 208–226. https://doi.org/10.1038/s41562-024-02059-4

Lin, C., & Thornton, M. (2023). Evidence for bidirectional causation between trait and mental state inferences. *Journal of Experimental Social Psychology*, *108*, Article 104495. https://doi.org/10.1016/j.jesp.2023.104495

Lin, Y., Michel, J.-B., Aiden, E. L., Orwant, J., Brockman, W., & Petrov, S. (2012). Syntactic annotations for the Google Books Ngram Corpus. In M. Zhang (Ed.), *Proceedings of the ACL 2012 System Demonstrations* (pp. 169–174). Association for Computational Linguistics. https://aclanthology.org/P12-3029/

Lindquist, K. A., Jackson, J. C., Leshin, J., Satpute, A. B., & Gendron, M. (2022). The cultural evolution of emotion. *Nature Reviews Psychology*, *1*(11), 669–681. https://doi.org/10.1038/s44159-022-00105-4

Lu, J. G., Song, L. L., & Zhang, L. D. (2025). Cultural tendencies in generative AI. *Nature Human Behaviour*, *9*(11), 2360–2369. https://doi.org/10.1038/s41562-025-02242-1

Mallinas, S. R., Crawford, J. T., & Cole, S. (2018). Political opposites do not attract: The effects of ideological dissimilarity on impression formation. *Journal of Social and Political Psychology*, *6*(1), 49–75. https://doi.org/10.5964/jspp.v6i1.747

Martin, A. E., Guevara Beltran, D., Koster, J., & Tracy, J. L. (2024). Is gender primacy universal? *Proceedings of the National Academy of Sciences of the United States of America*, *121*(35), Article e2401919121. https://doi.org/10.1073/pnas.2401919121

Martin, A. E., & North, M. S. (2022). Equality for (almost) all: Egalitarian advocacy predicts lower endorsement of sexism and racism, but not ageism. *Journal of Personality and Social Psychology*, *123*(2), 373–399. https://doi.org/10.1037/pspi0000262

Mattan, B. D., Kubota, J. T., & Cloutier, J. (2017). How social status shapes person perception and evaluation: A social neuroscience perspective. *Perspectives on Psychological Science*, *12*(3), 468–507. https://doi.org/10.1177/1745691616677828

McAdams, D. P. (1988). *Power, intimacy, and the life story: Personological inquiries into identity*. Guilford Press.

McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality*, *63*(3), 365–396. https://doi.org/10.1111/j.1467-6494.1995.tb00500.x

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist*, *61*(3), 204–217. https://doi.org/10.1037/0003-066X.61.3.204

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509–516. https://doi.org/10.1037/0003-066X.52.5.509

McCrae, R. R., Costa, P. T., Del Pilar, G. H., Rolland, J.-P., & Parker, W. D. (1998). Cross-cultural assessment of the five-factor model: The Revised NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, *29*(1), 171–188. https://doi.org/10.1177/0022022198291009

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Mohammad, S. M. (2025). *NRC VAD Lexicon v2: Norms for valence, arousal, and dominance for over 55k English terms*. arXiv. https://doi.org/10.48550/arXiv.2503.23547

Montiglio, P.-O., Ferrari, C., & Réale, D. (2013). Social niche specialization under constraints: Personality, social interactions and environmental heterogeneity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1618), Article 20120343. https://doi.org/10.1098/rstb.2012.0343

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, *56*(1), 63–75. https://doi.org/10.1177/0013164496056001004

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Nadler, E. O., Darragh-Ford, E., Desikan, B. S., Conaway, C., Chu, M., Hull, T., & Guilbeault, D. (2023). Divergences in color perception between deep neural networks and humans. *Cognition*, *241*, Article 105621. https://doi.org/10.1016/j.cognition.2023.105621

Neuberg, S. L., Williams, K. E. G., Sng, O., Pick, C. M., Neel, R., Krems, J. A., & Pirlott, A. G. (2020). Toward capturing the functional and nuanced nature of social stereotypes: An affordance management approach. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 62, pp. 245–304). Academic Press. https://doi.org/10.1016/bs.aesp.2020.04.004

Nicolas, G., Bai, X., & Fiske, S. T. (2022). A spontaneous stereotype content model: Taxonomy, properties, and prediction. *Journal of Personality and Social Psychology*, *123*(6), 1243–1263. https://doi.org/10.1037/pspa0000312

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, *66*(6), 574–583. https://doi.org/10.1037/h0040291

Nutter, S., Eggerichs, L. A., Nagpal, T. S., Ramos Salas, X., Chin Chea, C., Saiful, S., Ralston, J., Barata-Cavalcanti, O., Batz, C., Baur, L. A., Birney, S., Bryant, S., Buse, K., Cardel, M. I., Chugh, A., Cuevas, A., Farmer, M., Ibrahim, A., Kataria, I., … Yusop, S. (2024). Changing the global obesity narrative to recognize and reduce weight stigma: A position statement from the World Obesity Federation. *Obesity Reviews*, *25*(1), Article e13642. https://doi.org/10.1111/obr.13642

Oh, D., Shafir, E., & Todorov, A. (2020). Economic status cues from clothes affect perceived competence from faces. *Nature Human Behaviour*, *4*(3), 287–293. https://doi.org/10.1038/s41562-019-0782-4

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

OpenAI. (2024, May 13). *Hello GPT-4o*. https://openai.com/index/hello-gpt-4o/

Osgood, C. E. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist*, *66*(3), 171–200. https://doi.org/10.1525/aa.1964.66.3.02a00880

Peabody, D. (1987). Selecting representative trait adjectives. *Journal of Personality and Social Psychology*, *52*(1), 59–71. https://doi.org/10.1037/0022-3514.52.1.59

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology*, *57*(3), 552–567. https://doi.org/10.1037/0022-3514.57.3.552

Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLOS ONE*, *10*(10), Article e0137041. https://doi.org/10.1371/journal.pone.0137041

Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). W. W. Norton & Company. https://doi.org/10.1037/11494-000

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. [Technical report]. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjieh, R., Robertson, C. E., & Van Bavel, J. J. (2024). GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(34), Article e2308950121. https://doi.org/10.1073/pnas.2308950121

Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, *57*(4), 637–655. https://doi.org/10.1111/0022-4537.00233

Roberts, B. W., & Yoon, H. J. (2022). Personality psychology. *Annual Review of Psychology*, *73*(1), 489–516. https://doi.org/10.1146/annurev-psych-020821-114927

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*(4), 283–294. https://doi.org/10.1037/h0026086

Roy, E., Jaeger, B., Evans, A. M., Turetsky, K. M., O'Shea, B. A., Petersen, M. B., Singh, B., Correll, J., Zheng, D. Y., Brown, K. W., Kirgios, E. L., Chang, L. W., Chang, E. H., Steele, J. R., Sebastien, J., Sedgewick, J. R., Hackney, A., Cook, R., Yang, X., … Axt, J. R. (2025). A contest study to reduce attractiveness-based discrimination in social judgment. *Journal of Personality and Social Psychology*, *128*(3), 508–535. https://doi.org/10.1037/pspa0000414

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sainsbury, R. M. (1996). Concepts without boundaries. In R. Keefe & P. Smith (Eds.), *Vagueness: A reader* (pp. 251–264). MIT Press. https://philarchive.org/rec/SAICWB

Santos, H. C., Varnum, M. E. W., & Grossmann, I. (2017). Global increases in individualism. *Psychological Science*, *28*(9), 1228–1239. https://doi.org/10.1177/0956797617700622

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science*, *366*(6466), Article eaau5141. https://doi.org/10.1126/science.aau5141

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*(4), 689–699. https://doi.org/10.1037/0022-3514.52.4.689

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131–142. https://doi.org/10.1037/0033-2909.105.1.131

Smaldino, P. E., Lukaszewski, A., Von Rueden, C., & Gurven, M. (2019). Niche diversity can explain cross-cultural differences in personality structure. *Nature Human Behaviour*, *3*(12), 1276–1283. https://doi.org/10.1038/s41562-019-0730-3

Souza, F., Nogueira, R., & Lotufo, R. (2020). BERTimbau: Pretrained BERT models for Brazilian Portuguese. In R. Cerri & R. C. Prati (Eds.), *Intelligent systems* (pp. 403–417). Springer. https://doi.org/10.1007/978-3-030-61377-8_28

Stackhouse, M., Rickley, M., Liu, Y., & Taras, V. (2024). Homogeneity, heterogeneity, or independence? A multilevel exploration of Big Five personality traits and cultural values in 40 nations. *Personality and Individual Differences*, *230*, Article 112795. https://doi.org/10.1016/j.paid.2024.112795

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology*, *21*, 261–302. https://doi.org/10.1016/S0065-2601(08)60229-4

Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin*, *44*(4), 521–537. https://doi.org/10.1177/0146167217744194

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Thalmayer, A. G., Job, S., Shino, E. N., Robinson, S. L., & Saucier, G. (2021). ǂŪsigu: A mixed-method lexical study of character description in Khoekhoegowab. *Journal of Personality and Social Psychology*, *121*(6), 1258–1283. https://doi.org/10.1037/pspp0000372

Thalmayer, A. G., Mather, K. A., Saucier, G., Naudé, L., Florence, M., Adonis, T.-A., Shino, E. N., Asatsa, S., Witzlack-Makarevich, A.,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Bächlin, L. Z. M., & Condon, D. M. (2025). The cross-cultural big two: A culturally decentered theoretical and measurement model for personality traits. *Journal of Personality and Social Psychology*, *128*(5), 1181–1208. https://doi.org/10.1037/pspp0000528

Thalmayer, A. G., Saucier, G., Ole-Kotikash, L., & Payne, D. (2020). Personality structure in east and west Africa: Lexical studies of personality in Maa and Supyire-Senufo. *Journal of Personality and Social Psychology*, *119*(5), 1132–1152. https://doi.org/10.1037/pspp0000264

Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, *41*(1), 1–32. https://doi.org/10.1037/h0075959

Triandis, H. C. (2001). Individualism–collectivism and personality. *Journal of Personality*, *69*(6), 907–924. https://doi.org/10.1111/1467-6494.696169

Triandis, H. C., & Suh, E. M. (2002). Cultural influences on personality. *Annual Review of Psychology*, *53*(1), 133–160. https://doi.org/10.1146/annurev.psych.53.100901.135200

Unkelbach, C., Koch, A., & Alves, H. (2019). The evaluative information ecology: On the frequency and diversity of "good" and "bad". *European Review of Social Psychology*, *30*(1), 216–270. https://doi.org/10.1080/10463283.2019.1688474

Unkelbach, C., Koch, A., & Alves, H. (2021). Explaining negativity dominance without processing bias. *Trends in Cognitive Sciences*, *25*(6), 429–430. https://doi.org/10.1016/j.tics.2021.04.005

Varnum, M. E. W., Baumard, N., Atari, M., & Gray, K. (2024). Large language models based on historical text could offer informative tools for behavioral science. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(42), Article e2407639121. https://doi.org/10.1073/pnas.2407639121

Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*, *7*(3), 400–411. https://doi.org/10.1038/s42256-025-00986-z

Wang, X., Guinote, A., & Krumhuber, E. G. (2018). Dominance biases in the perception and memory for the faces of powerholders, with consequences for social inferences. *Journal of Experimental Social Psychology*, *78*, 23–33. https://doi.org/10.1016/j.jesp.2018.05.003

Williams, D. (2024, July 13). Ages of word types. *The Life of Words*. https://thelifeofwords.uwaterloo.ca/ages-of-word-types/

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, *113*(1), 59–80. https://doi.org/10.1037/pspi0000092

Yin, Z. (2018). *Ziyin-dl/ngram-word2vec* [Computer software]. GitHub. https://github.com/ziyin-dl/ngram-word2vec

Yzerbyt, V., Koch, A., Brambilla, M., Ellemers, N., Fiske, S. T., & Nicolas, G. (2025). Dimensions of stereotypes about groups. *Current Directions in Psychological Science*, *34*(6), 365–372. https://doi.org/10.1177/09637214251345489

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not competent but warm… really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, *8*(3), 291–308. https://doi.org/10.1177/1368430205053944

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, *26*(3), 237–242. https://doi.org/10.1177/0963721416683996

Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review*, *7*(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01

Received July 17, 2025
Revision received August 31, 2026
Accepted September 11, 2026 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.