Journal of Personality and Social Psychology

The Cross-Cultural Big Two: A Culturally Decentered Theoretical and Measurement Model for Personality Traits

Amber Gayle Thalmayer, Kendall A. Mather, Gerard Saucier, Luzelle Naudé, Maria Florence, Tracey-Ann Adonis, Elizabeth N. Shino, Stephen Asatsa, Alena Witzlack-Makarevich, Lea Z. M. Bächlin, and David M. Condon

Online First Publication, November 7, 2024. https://dx.doi.org/10.1037/pspp0000528

CITATION
Thalmayer, A. G., Mather, K. A., Saucier, G., Naudé, L., Florence, M., Adonis, T.-A., Shino, E. N., Asatsa, S., Witzlack-Makarevich, A., Bächlin, L. Z. M., & Condon, D. M. (2024). The cross-cultural big two: A culturally decentered theoretical and measurement model for personality traits.. *Journal of Personality and Social Psychology*. Advance online publication. https://dx.doi.org/10.1037/pspp0000528

AMERICAN PSYCHOLOGICAL ASSOCIATION

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

© 2024 The Author(s)
ISSN: 0022-3514

https://doi.org/10.1037/pspp0000528

The Cross-Cultural Big Two: A Culturally Decentered Theoretical and Measurement Model for Personality Traits

Amber Gayle Thalmayer[1, 2], Kendall A. Mather[3], Gerard Saucier[3], Luzelle Naudé[2], Maria Florence[4], Tracey-Ann Adonis[4], Elizabeth N. Shino[5], Stephen Asatsa[6], Alena Witzlack-Makarevich[7], Lea Z. M. Bächlin[1], and David M. Condon[3]

[1] Department of Psychology, University of Zürich
[2] Department of Psychology, University of the Free State
[3] Department of Psychology, University of Oregon
[4] Department of Psychology, University of the Western Cape
[5] Department of Psychology and Social Work, University of Namibia
[6] Department of Psychology, Catholic University of Eastern Africa
[7] Collaborative Research Centre (CRC 1252) "Prominence in Language", University of Cologne

A "big two" model has shown stronger cross-cultural replicability and links to theory than other contemporary models of personality trait structure. However, its theoretical and measurement models require better specification. We address this to create an initial English-language version of the Cross-Cultural Big Two Inventory with an empirically informed and culturally decentered approach, meaning that input from global contexts is used from the outset, without prioritizing Western perspectives. Four studies are reported: (1) Fifty-five items were identified from commonalities among 11 global lexical studies to define two factors. Communion/Social Self-Regulation captures the internalization of versus resistance to the normative codes of one's society, with components of warmth, morality, respect, industriousness, and even temper. Agency/Dynamism captures approach versus avoidance tendencies, with components of competence, confidence, fearlessness, positive mood, sociability, and surgency. (2) Items were reduced to the 45 most consistent across English-speaking contexts based on (a) frequency of use in World English corpora; (b) familiarity and exploratory factor analysis results among Africa Long Life Study participants, who were 18-year-olds from Namibia, Kenya, and South Africa ($N = 2{,}958$); and (c) distribution test statistics, exploratory factor analysis results, and test–retest reliability in online data from 13 diverse English-speaking countries ($N = 63{,}720$). (3) The 45-item Cross-Cultural Big Two Inventory was assessed psychometrically and validated against external criteria in the Africa Long Life Study samples and (4) in the online data and additionally compared to existing two-factor frameworks. The relation of the cross-cultural big two to other two-factor models and theories, its future development, and the potential and importance of culturally decentered models and inventories are discussed.

Keywords: cross-cultural assessment, big two, agency and communion, decolonization of psychology, personality structure

Supplemental materials: https://doi.org/10.1037/pspp0000528.supp

Kate C. McLean served as action editor.

Amber Gayle Thalmayer https://orcid.org/0000-0002-1963-1123
Kendall A. Mather https://orcid.org/0000-0001-7395-011X
Gerard Saucier https://orcid.org/0000-0003-3262-0469
Luzelle Naudé https://orcid.org/0000-0001-6660-8885
Maria Florence https://orcid.org/0000-0003-4854-7753
Tracey-Ann Adonis https://orcid.org/0000-0003-0610-0769
Elizabeth N. Shino https://orcid.org/0000-0002-1699-9635
Stephen Asatsa https://orcid.org/0000-0002-1045-2583
Alena Witzlack-Makarevich https://orcid.org/0000-0003-0138-4635
Lea Z. M. Bächlin https://orcid.org/0009-0002-8667-0402
David M. Condon https://orcid.org/0000-0002-8406-783X

All data, analysis code, research materials, and the study design and analysis were preregistered at the Open Science Framework available at https://osf.io/89x2q/.

Africa Long Life Study data collection was supported in Namibia by Selma Uugwanga, Arthur Kambambi, Elveraldo Awases, Ester Muhepa, Grace Mubiana, Hilma Awala, Jasmine Hangara, Julius Ndikwetepo, Laina Shangano, Lorensia Kafuro, Marialda Garises, Monika Matakala, Paulus Mwetulundila, Regina Mwanyangapo, and Taimi Amutse. In Kenya, it was supported by Daniel Hofmann, Wairimu Murichi, John Makunda, Anna Madoka, Alice Sangok, Esther Kerebi Nyamanche, Eucabeth Manyibe, Esther Wangu, Godwin Emochi, Lorna Ayako, Lawrian Mseri, Muthoni Mwangi, Margaret Natocho, Mikhal Odhiambo, Mercy Wambui Wachira, Monica Nyagah, and Sammy Mutisya. In Central South Africa, it was supported by Elam Maso, Kutloano Molumo, Kgabo Mphela, Boitumelo Sehlabaka, Lerato Sokhaya, and Zandile Viti and on the Western Cape by Casey Botha, Adhelle van der Poll, Amy-Jean Viljoen, Ashton Smith, Erika Erasmus, Jana Ras, Khulekani Magagula, Leano Cleophas, Koketso Nelly Mamabolo, Nnenna Osonda, Noluthando Mpisane, Precious Ndebele, Safaa Reddy, Sinesipho Jenkins, Siphiwe Kobo, Tashnikah Davids, Valentina Bantam, Yonela Ngubo, Shahied Visagie, Hannah Lukic, Khelsey Fraser, Xolisa Gwadiso, Munene Nkuna, Shannen Ferreira, and Warona Mateane. Data processing additionally relied on Daniel Hofmann and Julia Rotzinger. Christopher Hopwood helped create Figure 2.

Funding for the preparation of the article and for data collection in the Africa Long Life Study was provided by the Swiss National Science Foundation Eccellenza fellowship to Amber Gayle Thalmayer, "Universal Specific Versus Culturally-Specific Aspects of Personality and Mental Health," Grant

continued

Personality psychologists have long recognized the utility of a cross-culturally relevant theory and structural model for personality traits (e.g., L. Goldberg, 1981). After early efforts using lexical studies of personality, among other means, the Big Five was hailed as a potentially universal trait model (McCrae & Costa, 1997). Recent evidence suggests that while the Big Five is a useful, parsimonious model that has been highly generative for personality psychology, it does not meet optimal standards for cross-cultural universality (Cheung et al., 2013; Laajaj et al., 2019; Thalmayer et al., 2022). While the Big Five still serves its role of predicting behavior in Western contexts, the need remains for a theoretical and taxonomic model that is culturally decentered, meaning that it should be developed with data from populations around the world without prioritizing a single context or kind of society. Such a model could facilitate cross-cultural comparison, help disambiguate more ubiquitous from more culturally specific traits, and provide a basis for theory where cross-culturally universal traits are relevant, for example with regard to evolutionary psychology.

A model of two broad factors may fit the bill. Such a model has been shown to arise spontaneously in highly diverse indigenous lexical studies (Saucier, Thalmayer, & Bel-Bahar, 2014), indicating its potential as a culturally decentered model, with compelling links to theoretical two-factor models (Bakan, 1966; R. Hogan & Cheek, 1983; Wiggins, 1979). The big two potentially provides a taxonomy at a higher level, under which higher dimensional, locally useful traits and models, such as the Big Five, might be organized (DeYoung, 2006; Digman, 1997; Irwing et al., 2024; Saucier, Thalmayer, & Bel-Bahar, 2014). This approach could allow for more diversity in specific traits and facets, helping to emancipate personality psychology from the limits of the Big Five (e.g., Mõttus et al., 2020; Saucier & Iurino, 2020) while integrating Big Five findings. In the study of both emotions and psychopathology, consideration is given to both universal and culturally specific constructs, with those having the highest cross-cultural ubiquity prioritized when considering underlying biological models. Personality psychology could benefit from a similar approach. This would be facilitated by a theoretical and structural model of the most ubiquitous traits.

The aim of the studies reported here was to systematically integrate cross-cultural evidence to (a) refine a big two theoretical model as a foundational taxonomic structure for personality trait variation with strong links to theoretical bases, which can serve as an organizational rubric for higher dimensional traits, including those common across many societies and those that are more culture-specific, and (b) to develop and validate an associated assessment model—an empirically informed English-language Cross-Cultural Big Two Inventory (CCB2I)—that can be used across global English-speaking contexts and is optimized for translation into other languages. Additionally, this project was pursued and is presented as a creative attempt at cross-cultural theory and scale development, making use of evidence from around the globe from the outset, rather than developing a model in the West and then importing it to other contexts. Following a brief review of the empirical and theoretical background for a big two model, the culturally decentered strategy is described in detail, with details about the four studies reported here.

Lexical Studies and the Big Two

Lexical studies of personality traits have been central in developing structural models of personality traits in the last century (Thalmayer et al., 2022). They are based on the rationale from linguistics that concepts expressed often will become encoded into a single word in the language. Thus, the more relevant a trait is to a community of speakers, the more words for it can be found in their lexicon (Allport & Odbert, 1936), and the more important a concept is for humans in general, the more languages should have a word for it (L. Goldberg, 1981). While cross-cultural ubiquity is certainly not the only criteria by which a personality model can be judged, this standard is an excellent starting point for enhancing objectivity, especially cross-cultural objectivity.

Lexical studies are conducted in four basic steps, which are transferable to any language with a comprehensive written lexicon. (a) All terms used to describe psychological differences are extracted from a comprehensive lexicon. (b) The list is reduced to a tractable number by removing redundant and unfamiliar terms. (c) The list is administered to participants, asking them how well each term describes a target

PCEFP1_194552. Additional funding for Wave 1 data collection in South Africa came from the Southern African Medical Research Council (University of the Western Cape) awarded to Maria Florence and the psychology departments of the University of the Free State and the University of the Witwatersrand. This research was funded in whole, or in part, by the Swiss National Science Foundation (SNSF) (Grant PCEFP1_194552). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

This work is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.0; https://creativecommons.org/licenses/by-nc-nd/4.0). This license permits copying and redistributing the work in any medium or format for noncommercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Amber Gayle Thalmayer played a lead role in project administration, supervision, writing–original draft, and writing–review and editing and an equal role in conceptualization, funding acquisition, and methodology. Kendall A. Mather played a supporting role in conceptualization, formal analysis, methodology, and writing–review and editing. Gerard Saucier played a supporting role in methodology and writing–review and editing and an equal role in conceptualization. Luzelle Naudé played a supporting role in data curation, funding acquisition, project administration, resources, supervision, and writing–review and editing. Maria Florence played a supporting role in funding acquisition, project administration, supervision, and writing–review and editing. Tracey-Ann Adonis played a supporting role in data curation, supervision, and writing–review and editing. Elizabeth N. Shino played a supporting role in project administration, resources, supervision, and writing–review and editing. Stephen Asatsa played a supporting role in data curation, investigation, project administration, resources, supervision, and writing–review and editing. Alena Witzlack-Makarevich played a supporting role in data curation and writing–review and editing and an equal role in conceptualization, formal analysis, and methodology. Lea Z. M. Bächlin played a supporting role in data curation, formal analysis, and writing–review and editing. David M. Condon played a supporting role in project administration, resources, and writing–review and editing and an equal role in conceptualization, funding acquisition, data curation, formal analysis, and methodology.

Correspondence concerning this article should be addressed to Amber Gayle Thalmayer, Department of Psychology, University of Zürich, Binzmühlestrasse 14, Box 19, Zürich 8050, Switzerland. Email: ambergayle.thalmayer@uzh.ch

(the self or a well-known other). (d) Factor analysis is used to determine which terms group together into traits and which best distinguish between individuals in the population (Thalmayer et al., 2022). In a separate step, additional terms can be included as needed to create marker scales for imported models, which can be correlated with emic factors for empirical tests of replication (e.g., Thalmayer, Job, et al., 2021).

Qualitatively comparing the factors identified in the lexicons of different languages and quantitatively comparing their replicability using marker terms can lead to establishing a cross-cultural model. The earliest lexical work focused on the Big Five, but the familiar five dimensions do not readily emerge in many cultures and languages (De Raad et al., 2014; Saucier, Thalmayer, Payne, et al., 2014). Overall, the two dozen or more indigenous lexical studies so far conducted around the globe do not support a Big Five structure as ubiquitous across cultural contexts (summarized in detail by Thalmayer et al., 2022). While Big Five inventories have been translated and exported for use all over the world (Allik et al., 2017; Bleidorn et al., 2013; McCrae & Costa, 1997; Schmitt et al., 2007; Van Aarde et al., 2017) even with forager–farmers in the Amazon (Gurven et al., 2013), they tend toward low reliability outside the West (e.g., Laajaj et al., 2019) and have never been supported by a level of measurement invariance which would allow for valid comparison of scores across contexts (Church et al., 2011; Rossier et al., 2017; Thalmayer & Saucier, 2014).

Alternative models have been proposed, for example, the HEXACO and Big Six, which are based on more diverse sets of lexical studies and are typically more orthogonal to one another (Ashton et al., 2004; Saucier, 2009), but these have not fared better than the Big Five in terms of measurement properties in large-scale international studies (Thalmayer & Saucier, 2014; Thielmann et al., 2020). De Raad et al. (2014) used Procrustean factor techniques with lexical study data to identify the pan-cultural three model, which, however, is based almost exclusively on European languages. One-factor structures are likely contextually invariant, given the universality of good versus bad moral evaluation (e.g., Thalmayer, Saucier, et al., 2021; Wierzbicka, 1996), but this is a sparse model; if a model with more factors were similarly universal, it would clearly be preferable.

Regularities in two-factor models of global lexical studies provide a potential solution. The content of the first two factors from nine globally diverse lexical studies, derived when hundreds of natural-language descriptors, translated into English as reported by the original study authors, converge well on a big two model (Saucier, Thalmayer, Payne, et al. 2014). Importantly, this approach is based on data from diverse populations around the world. The cross-cultural replicability of this model was supported in three African lexical studies (in Maasai and Senufo, Thalmayer et al., 2020; and in Khoekhoe, Thalmayer, Job, et al., 2021), unlike Big Five and Big Six/HEXACO models. For this reason, the results of Saucier, Thalmayer, Payne, et al. (2014) are used here as a starting point for the development of a cross-cultural big two theory and inventory. These refine and update earlier "big two" theories in terms of their relevance across human societies.

The Theoretical Background for a Big Two Model

Converging with the regularities in two-factor models of global lexical studies (Saucier, Thalmayer, Payne, et al., 2014), many two-factor nomothetic models have been proposed, based on methods outside the lexical approach and in other subdisciplines of psychology, including clinical, organizational, and psychobiological systems theory. Additionally, empirical evidence from survey studies of higher order dimensions of the Big Five traits and the structure of psychopathology provide support for the utility of two basic domains of individual differences. In the following subsections, we review these models and their relations to the work proposed here.

Agency and Communion

Two-factor models have been posited in many Western frameworks, now generally subsumed under the terms Agency and Communion. Adler (1938) defined psychological dimensions of social interest and superiority striving. Bakan (1966) described Agency and Communion as competing psychological trends in the history of Western thought and religion. Wiggins' (1979) used the concepts of nurturance and dominance as the basis for a circumplex inventory (Wiggins & Broughton, 1991). R. Hogan (1982) described conformity and sociability, or getting along and getting ahead, as accounting for the major individual differences in vocational and social roles, with other traits describing how well individuals perform within those roles. McAdams (1988) described power and intimacy motives, capturing degrees of preference for personal impact and for interpersonal closeness (Zeldow et al., 1988).

In studies of stereotypes, person perception and judgment, and impression formation, social psychologists have relied on dimensions of competence and of morality (Wojciszke, 2005) or warmth (Fiske et al., 2007; Judd et al., 2005), later termed Agency and Communion, following Bakan (1966; Abele et al., 2016; Abele & Wojciszke, 2014; Wojciszke et al., 2009). Brewer and Chen (2007) defined these dimensions as capturing fundamental human needs for uniqueness and for social belonging. Similar motivational bases are captured by the most frequently used dimensions of cross-cultural comparison: individualism (Markus & Kitayama, 1991; Wojciszke, 1997, 2005), concerning differentiation from others and the ambient culture (Gebauer et al., 2013), and collectivism, concerning assimilation to others and the ambient culture (Sedikides et al., 2003). Underlying these and other models (e.g., Frimer et al., 2011; Ybarra et al., 2001) are two motivational bases: to assert and expand the self and to integrate the self in a larger social unit (Gebauer et al., 2013). Drawing on evidence and theory from across domains of psychology, contemporary integrative interpersonal theory shows how the dimensions of agency and communion can be used to integrate models of personality traits and of psychopathology (Hopwood et al., 2021; Pincus et al., 2020).

Higher Order Factors of the Big Five

Digman (1997) found consistent patterns of correlation among Big Five survey results, with higher order factors that approximately combine Agreeableness, Conscientiousness, and Emotional Stability ("α," conceptually close to Communion), and Extraversion and Openness/Intellect ("β," conceptually close to Agency). DeYoung (2002, 2006) found the same patterns in self- and informant reports in a large community sample and proposed the labels stability, for the tendency to avoid disruption (low Neuroticism, high Agreeableness and Conscientiousness), and plasticity, for the tendency to explore and engage flexibly with novelty (aspects of Extraversion and Openness). Jang et al. (2006) found substantial heritability and a

genetic source for the coherence of the two higher order factors of the Big Five in a sample of nearly 2,000 twin pairs in Canada, Germany, and Japan.

The Structure of Psychopathology

While psychopathology is a narrower component of individual differences than personality traits, there may be useful correspondences. Common psychological disorder symptoms, based on comorbidity patterns and symptom checklists, have been shown to consistently aggregate into overarching spectra (Conway et al., 2019; Kotov et al., 2017). The primary two spectra include internalizing, which captures fear-based withdrawal from the external world (anxious, traumatic, phobic, depressed, obsessive, and compulsive symptoms) and thus relates logically to low Agency. Externalizing captures a tendency to move against the world, in conflict with social and moral norms (aggressive, impulsive, delinquent, and substance-use symptoms) and thus relates to low Communion. Individual differences in these factors have high stability (Krueger et al., 1998; Vollebergh et al., 2001) and strong genetic underpinnings (Kendler et al., 2003, 2011; Krueger et al., 2002; Lahey et al., 2017).

An Integrated Cross-Cultural Big Two Theory

These various two-factor models and sources of evidence for two basic factors of human variation stem from a variety of approaches and data sources. They share an intriguing degree of overlap but may also differ. Moreover, a limitation of this prior work is being primarily based in Western data, which could lead it to unintentionally reflect cultural biases. The current project thus takes as its starting point the convergent patterns found in languages of diverse provenance including Asia and Africa as well as Europe (Saucier, Thalmayer, Payne, et al., 2014), creating a lexical foundation for the big two that is not derivative of prior Big Five work or other theories. The current model is derived empirically, with input from diverse cultures, in order to build on a more generalizable foundation, which can be systematically compared to prior models to assess for consistency or divergence, and to refine an integrated and cross-cultural big two theory.

What does the evidence suggest about the two factors? Personality terms, as translated into English by the authors of nine lexical studies, were compared by Saucier, Thalmayer, Payne, et al. (2014). Those concepts that appeared in the majority of the nine languages, in what was the first factor in seven studies, were: *honest*, *kind*, *generous*, *gentle*, *good*, *obedient*, *respectful*, *diligent*, *responsible*, and *(−)selfish*. The label Social Self-Regulation (SSR) was chosen by those authors to convey how this domain relates to the internalization of social and cultural norms, as it was seen to go beyond the conceptions conveyed by the labels communion, nurturance, union, or warmth, all of which focus on what could be conceptualized as Big Five Agreeableness. The cross-culturally derived dimension additionally includes content related to honesty/integrity and to Big Five Conscientiousness. Moral evaluation would also be an apt and appropriately broad term.

The second factor (this content came first in only two studies) was defined by the terms *active*, *brave*, *bold*, *lively*, *(−)timid*, *(−)weak*, and *(−)shy*. This set of descriptors describes relative proportion of approach versus avoidant tendencies, as well as social effects from interesting to boring (as in Saucier, 2010). This dimension was labeled Dynamism by Saucier, Thalmayer, Payne, et al. (2014), following Tanaka and Osgood (1965), who used it for a blend of activity and potency in description.

Saucier, Thalmayer, Payne, et al. (2014) constructed scales with International Personality Item Pool (IPIP) terms in a U.S.-based community sample to measures their cross-cultural big two and compare it to other two-factor models including the interpersonal circumplex, the stereotype content model, the Wojciszke model, internalizing and externalizing tendencies, behavioral activation versus inhibition, and reward versus punishment sensitivity, as well as Big Five and Six scales. SSR was closely associated with Communion scales ($r_{\text{Nurturance}} = .65$, $r_{\text{Warmth/Morality}} = .68$, $r_{\text{Morality/Communion}} = .65$), externalizing tendencies in terms of self-report psychopathy ($r = -.43$), and Agreeableness in Big Five ($r = .52$–$.75$) but not Big Six models ($r = .28$–$.30$) and had virtually no association with behavioral activation or inhibition. Associated items suggest that the domain concerns how much a person adheres to rules, behaves "properly," shows respect for others and authority, and keeps promises. Each of these relates to using societal norms and rule systems as standards for regulating behavior. Dynamism associated with Agency scales ($r_{\text{Dominance}} = .73$, $r_{\text{Competence}} = .52$, $r_{\text{Competence/Agency}} = .53$), internalizing tendencies in terms of sensitivity to punishment ($r = -.56$), and the ratio of activation/reward to inhibition/punishment sensitivity ($r = .54$), and Extraversion ($r = .45$–$.78$).

These results can be summarized into theoretical postulates for how the big two capture fundamental driving factors in human variation: (a) the degree to which an individual internalizes versus resists the normative codes of their society and (b) an individual's setting with respect to taking risks to exploit opportunities versus playing it safe to promote survival, that is, how approach are weighed against avoidance motives. Other aspects of between-person variation are posited to vary more between contexts. Similar to Hogan's socio-analytic theory and Pincus and colleagues' contemporary integrative interpersonal theory, we argue that it is not so much biology as the social–adaptive context that creates these lines of variation. Humans learn both normative and strategic codes (Bailey, 2018), one oriented toward harmonious coordination with others, the other to achievement of personal goals; the first is more connected with SSR and moral evaluation, the latter with Dynamism. We must decide how and where to prioritize one with respect to the other, and we tend to resolve these recurrent dilemmas in a somewhat characteristic, individualized way.

In summary, a cross-cultural big two model is proposed to serve personality psychology by integrating and refining existing theoretical models for psychological variation and providing a culturally decentered basis for the most universal dimensions. Higher dimensionality, locally relevant structures such as the Big Five might be organized within this overarching structure. The next step is to establish an assessment model for the cross-cultural big two that works well across contexts, the data from which can be used to further refine the theoretical model for global human psychology. Current Agency and Communion measures (e.g., Abele et al., 2016; Buchanan & Bardi, 2015; Gebauer et al., 2013; Trapnell & Paulhus, 2012) and the Big Five or Six measures, which could be scored for higher order factors, do not reflect the same content as in Saucier, Thalmayer, Payne, et al. (2014). They were developed based on evidence nearly exclusively from the West and have not been supported with compelling cross-cultural evidence. An ideal inventory for a cross-cultural model would be developed with cross-cultural evidence at every stage and decision point.

We acknowledge from the outset that "big few" models have inherent limitations when it comes to predicting specific behaviors

and outcomes. Even the Big Five are overly broad for many purposes (e.g., Mõttus, 2016), and a two-factor model is even broader. Scores on two factors can be expected to associate moderately with many behaviors and outcomes, without providing great specificity. This means that for many purposes, more fine-grained models and inventories are needed. Big two traits, however, are highly intuitive—they fit how people actually perceive and speak about others, likely across global contexts (e.g., Wojciszke, 2005). A big two model offers the potential for a solid basis of cross-cultural comparison in personality description and research around the globe and for a higher level taxonomy under which useful fine-grained levels of society-specific traits can be organized. This could be a starting point for efforts to better disambiguate more ubiquitous from more culturally specific traits and for theory building where cross-culturally universal traits are relevant.

Culturally Decentered Model and Inventory Development

Standard practice in psychology has been to develop, refine, and validate an inventory in the West; only after it is finalized is it translated into other languages and assessed for cross-cultural applicability. This is problematic for several reasons. First, it implies that Western populations set the norms for all other populations. Second, as the inventory is already published and in widespread use when exported, updates or refinements based on fit in other contexts are unwelcome. This creates a strong incentive to find fit and a disincentive to seriously explore divergences. Further, it discourages the integration of new information from other contexts despite increasing awareness of the peculiarities of psychology in Western populations (e.g., Henrich, 2020). This limits potential to improve the generalizability of inventories and to distinguish more universal from more culturally specific aspects of psychology. Concerningly, it strengthens a pattern of colonialism and disrespect in science, giving the message that divergences from the West are wrong, or at least abnormal (e.g., Adams et al., 2015; Silan et al., 2021).

One example of treating the United States as the normative reference point in the NEO-PI-R, which has been translated into at least 40 languages (Costa & McCrae, 2014), is an item asking whether one would enjoy a trip to Las Vegas. Most people in the world are unfamiliar with the significance of "Vegas," and many regions lack an analogous city or even a tradition of tourism or gambling.

With the present project, we sought to develop a more "culture-fair" approach to cross-cultural modeling and inventory development. Perfectly culturally decentered models and inventories in psychology may represent an unattainable ideal, but we can make creative efforts to develop them in ways that are relatively culturally decentered. Our first aim being the development and refinement of an English-language inventory, our initial interest is in global representation of English-speaking countries. We explain below the practical reasons for a focus on English, the ways that an English-language survey can and should include global input, and the ways this project leaves the door open for future refinements once cross-language evidence becomes available.

Why English?

The practical reasons to begin this project in English are threefold. First, the initial model is derived from the results of 11 lexical studies that provide global representation. While these studies were conducted in 11 languages, comparison requires a common language, and all were published in English, which means that the terms in the studies were published with English translations. These translations form the basis for comparison of content, leading to the initial item pool in English.

Second, English is the most-spoken language in the world and the official or de facto language in 63 countries (Central Intelligence Agency, 2021). The way English initially spread via British colonialism included violence and erasure of cultural and linguistic diversity. Today, however, English has been self-chosen as a lingua franca in many multilingual nations, where it develops into local varieties suited to contextual needs. This makes English a potentially democratic choice for a survey instrument, so long as the experiences of international speakers and differences in usage across contexts are taken into account. The countless forms of English now in use are collectively referred to as "World Englishes" (Melchers et al., 2019). In addition to "traditional" varieties from within Great Britain, new varieties have emerged that range from the institutionalized (Kirkpatrick, 2012), supraregional, and internationally recognized to the informal and localized, including contact-derived pidgins and creoles and second-language learner varieties (Schreier et al., 2020). The many differences between World Englishes pertain to phonology and morpho-syntax (Kortmann & Schneider, 2004a, 2004b) and, most importantly for our purposes, to their lexicons, that is, the presence, meanings, and usage of individual terms. These differences are rarely addressed in cross-cultural research, but we integrate this concern from the outset. This creates a rare instance of transporting a model "northward": It is developed with data from (largely) majority-world contexts then tested in Western contexts.

Finally, a major impetus for the development of this inventory was the need for a valid and reliable personality trait inventory for use in sub-Saharan Africa. The majority of coauthors are based at English-speaking African universities and/or associated with the Africa Long Life Study (ALLS), a three-country longitudinal study of African young adults described further in Studies 2 and 3. The low reliability and validity of Western inventories in these contexts make it difficult to pursue personality research in the region. While some local inventories have been created (e.g., Fetvadjiev et al., 2015; Thalmayer, Saucier, et al., 2021), they are specific to certain languages or contexts and not suitable for the multicountry ALLS, which includes speakers of over 50 local languages (Thalmayer et al., 2024). In all three countries, English is one of the nationally recognized languages and is used predominantly in secondary and higher education. A personality inventory in English, using a cross-cultural theory, model, and items, was needed for the ALLS to effectively test questions related to personality psychology. Important in these and many other global contexts, however, is to take seriously the impact of multilingualism and second-language status and to avoid unnecessarily complex or unfamiliar phrasing and terminology.

This initial effort forms a basis for translations to be created in future studies. Additional refinements to the inventory can and should be made to the CCB2I as soon as cross-language evidence becomes available. For this reason, the inventory will remain as long as practically possible at this stage so that a larger pool of terms is available to translators and for researchers across contexts to draw from.

Goals for the Present Studies

As described above, we carefully considered how to create a model and an inventory with a culturally decentered approach, focusing on World Englishes from five world regions (Africa, Asia, Latin America/Caribbean, Europe, and North America) to form a robust cross-cultural basis for later translation. We hope this will inspire other teams to imagine ways to create cross-culturally optimized, culturally decentered models and measures.

In Study 1, we expand on the nine languages in Saucier, Thalmayer, Payne, et al. (2014), to develop a preliminary item pool for measuring the cross-cultural big two using recurrent concepts from 11 globally representative lexical studies. The content of (and labels for) the cross-cultural big two theoretical model are discussed in relation to other big two models.

In Study 2, we refine the pool by assessing the suitability of the items across English-speaking contexts in multiple ways: In 10 corpora of World Englishes, we consider the presence versus the absence of the terms. In ALLS samples of 18-year-old participants from three African countries, we (a) identify terms that were rated as unknown and (b) explore their factor structure; in data collected online in 13 English-speaking countries (the Synthetic Aperture Personality Assessment [SAPA] Project; https://sapa-project.org), we consider (a) distribution test statistics (DTS) to evaluate similarity of responses, (b) factor structure, and (c) within-session test–retest reliability. These sources of information are integrated to identify the most and least promising items to refine the CCB2I.

In Study 3, the psychometric properties and measurement invariance of the refined CCB2I and the scales' association with external criteria, including mental and physical health, gender, and religious engagement, are tested in the ALLS data. In Study 4, the CCB2I scales' psychometric properties, validity against external criteria, and association with other two-factor frameworks are tested in the 13-country SAPA data, allowing for an in-depth consideration of how the cross-cultural big two overlap with related models. Together, these four studies provide evidence from global contexts to refine a theoretical framework and shape a culturally decentered English-language model and inventory. They also present and explore approaches that might be used to shape a culturally decentered personality science for the future.

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, we and follow Journal Article Reporting Standards for quantitative research in psychology (Appelbaum et al., 2018). All data, analysis code, research materials, and the study design and analysis were preregistered at the Open Science Framework (OSF) at https://osf.io/89x2q/. Data were analyzed using R, Version 4.2.1 (R Core Team, 2022), with respective packages cited for each study component.

Study 1: A Culturally Decentered Item Pool for Measuring the Big Two

Study 1: Method

Recurrent items were identified for two factors from 11 original lexical studies. This included the original nine of Saucier, Thalmayer, Payne, et al. (2014; Chinese, Korean, Filipino, Turkish, Greek, Polish, Hungarian, Maasai, and Senufo) and additionally Malay (Mastor et al., n.d.) and Khoekhoe (Thalmayer, Saucier, et al., 2021). The top 50 terms from each factor of a two-factor solution obtained using varimax rotation, provided by the original study authors, were obtained for each language, leading to a list of 550 total terms, each paired with an English-language "gloss" (translation). A search was made by English word root among the list of glosses, to identify where the same English concept appeared across languages. A relatively relaxed criterion of appearing in at least four of 11 languages (as opposed to only those appearing in five or six) was chosen at this stage in order to provide a maximally long initial list and to account for idiosyncrasies in translation (i.e., in some cases a less common English term may have been chosen by translators instead of its more common synonym).

Study 1: Results

A total of 55 terms, in English-translation, appeared in at least four of the 11 language lists. These included 30 for the first factor (Communion/SSR) and 25 for the second (Agency/Dynamism). These terms are displayed in Table 1, sorted with near synonyms on each row, and by content clusters. Some clusters will be familiar from other big two models (warmth, morality, competence), others less so (surgency). The languages in which each term was found are detailed in Supplemental Table S1.

Study 1: Discussion

The goal of Study 1 was to build on Saucier, Thalmayer, Payne, et al. (2014) to develop a preliminary item pool for measuring the cross-cultural big two by identifying key concepts/terms in a sample of 11 globally representative lexical studies. The content and the labels for the dimensions were also considered carefully. The terms Agency and Communion have been increasingly employed in psychology, displacing many of the other terms previously used, and were deemed "the big two" by Gebauer et al. (2013).

The 55 terms identified are clustered by content in Table 1 for comparison to other models and inventories. The first dimension, Communion/SSR, is similar to that defined by many big two inventories in that it includes components of warmth and morality, but it also includes content that goes beyond prior inventories (e.g., Abele et al., 2008, 2016; Abele & Hauke, 2020; Buchanan & Bardi, 2015; Fiske et al., 2002; Gebauer et al., 2013; Trapnell & Paulhus, 2012; Wojciszke et al., 2009). Most notably, it includes terms related to Big Five Conscientiousness (*diligent*, *industrious*, *hardworking*) and Emotional Stability, specifically, "even temper" (*calm*, *peaceful*, *hot-tempered*, *hotheaded*), which make it closer to the empirically identified higher order factor of the Big Five (DeYoung, 2006; Digman, 1997). Given the breadth and nature of this content, Saucier, Thalmayer, Payne, et al.'s (2014) SSR label is arguably a more accurate fit than the more popular Communion. However, they are both acceptable, as long as one takes a broader view of Communion-related traits than is reflected in the inventories cited above. We use the two labels together in what follows to be clear that refining and broadening the concept is necessary.

The second dimension includes a significant component of the competence content that prior inventories emphasize or exclusively measure (Fiske et al., 2002; Gebauer et al., 2013; Wojciszke et al., 2009). Our model, however, includes equally significant components

Table 1
Preliminary Item Pool for the Cross-Cultural Big Two, Grouped Loosely by Content

Communion/Social Self-Regulation	Agency/Dynamism
Warmth	Competence
Benevolent	Clever
Compassionate	(−) weak
Generous	Confidence
Kind, **kindhearted**	Confident, **self-confident**
Good-natured	(−) timid
Friendly	Fearlessness
Morality	Brave, courageous, (−) **cowardly**
humane, (−) inhumane	Bold, daring
(−) bad	Determined
Honest	Positive mood
Faithful	Happy, (−) unhappy, (−) sad,
Respect	Cheerful
Courteous, polite	(−) pessimistic
Modest, (−) arrogant	Sociability
Respectful	(−) a loner
(−) gossipy/gossip[a]	(−) **bashful**, (−) shy
Obedient, (−) disobedient	Surgency
Industriousness	Active, dynamic, energetic
Diligent, **industrious**, hardworking	(−) quiet, (−) silent
Even Temper	Lively
Calm, peaceful, (−) hot-tempered, (−) **hotheaded**	
Gentle, (−) aggressive	

Note. Items ultimately chosen for removal, based on Study 2 results, are in bold. Terms on the same line are close synonyms or antonyms. The term *naïve* emerged in the results for Communion but is excluded from the table as it did not fit into any topic groupings.
[a] Changed in form after Study 2, from adjective to verb "I tend to gossip.".

related to surgency/energy level and to mood state. This splitting of Emotional Stability between the two dimensions distinguishes the CCB2 from higher order factors of the Big Five (DeYoung, 2006; Digman, 1997), which treat the Big Five traits as single entities and thus cannot incorporate this kind of important nuance. Absent from our second factor is content related to dominance, which is prominent in some theories of Agency (Abele et al., 2008; Trapnell & Paulhus, 2012). While the term Agency arguably implies more intentionality and potency, it is similar in meaning to the term Dynamism from Saucier, Thalmayer, Payne, et al. (2014), and both fit the observed content well.

A limitation in this study is that we used a low standard of universality, requiring a term to only appear in four of 11 languages to be added to our list. This was to achieve a larger list and to account for "noise" in the data set that is introduced by idiosyncrasies of translation. A related limitation was our dependence on translations published by study authors. Following from these, we have a list that includes some relatively unfamiliar terms, for example, *bashful*, *benevolent*, and *humane*, as well as synonyms using the same word root (e.g., *kind-hearted* and *kind* both met criteria to be included). We also have one word that could not be framed as an adjective (*a loner*) and another that was framed as an adjective but is rarely used this way (*gossipy*). These issues flag some terms as candidates for removal from the outset, though they were retained into Study 2. Furthermore, these lumpy and complex factors almost certainly lack unidimensionality (i.e., a pattern of relatively even interitem correlations); they were created following an empirical process focused on identifying content rather than optimizing psychometric properties. The subsequent studies inform us empirically which terms to retain.

Study 2: Refinement of CCB2 Items With Exploration Across English-Speaking Contexts

Study 2 includes three parts, each providing information to assess how the 55 items work across global English-speaking contexts. The goal was to identify items that capture the two separate components of the model and that are used and understood (in terms of valence, familiarity, and associations) most consistently and similarly across contexts. At the same time, we sought to identify the most problematic items across a variety of criteria, which may require removal due to poor or differential performance across contexts.

Study 2.1: Frequency of Use in 10 International Corpora of English

As noted above, the widespread use of English has given rise to processes of linguistic diversification that are unparalleled on a global scale. Discussing the lexical differences between Standard American and British English, Kretzschmar (2010) notes that speakers of these varieties are familiar with a small number of typical differences, such as *traffic circle* versus *roundabout*, but most differences can only be discovered in terms of statistical preferences in corpus-linguistic studies. Using corpora from 10 varieties/dialects of English from the International Corpus of English (ICE, http://ice-corpora.net/ice/), we assess the frequency with which our 55 terms appear across corpora of World Englishes on the grounds that terms that are only used in some global contexts (e.g., the United States) but not in others (e.g., English-speaking countries in Africa, Asia, or the Caribbean) should be considered for removal from the CCB2I.

Study 2.1: Method

Access was obtained to corpora of as many varieties of World Englishes as possible, amounting ultimately to 10: ICE-Nigeria, ICE-East Africa (which includes texts from Kenya and Tanzania), ICE-Hong Kong, ICE-Philippines, ICE-India, ICE-Sri Lanka, ICE-Jamaica, ICE-United States, ICE-Canada, and SPICE-Ireland. All details including citations and code are provided in additional online materials "Corpora Analysis" available at https://osf.io/89x2q/. As corpora can differ in size, it is common to use normalized instead of raw frequencies, for example, taking a normalization base of appearances per 1 million words. Our focus here, though, was simply on terms with the raw frequency of 0.

Study 2.1: Results and Discussion

The number of corpora from which each item was missing were counted and are listed in order of most frequently missing in the first column of Table 2. The terms *bashful*, *good-natured*, and *hotheaded* did not appear in any corpora. *Hot-tempered*, *loner*, and *kind-hearted* were missing from all but one or two corpora, and *gossipy* from seven. These terms may be little used outside the United States, even by those whose first language is English. In fact, the first three terms, while familiar to many Americans, were missing in the ICE-United States corpus, indicating that these are not commonly used in the United States either (i.e., their frequency is below once per million words in ICE-United States).

The absence of a term from a corpus does not mean that it is entirely unknown in a variety of English but, rather, that it is quite infrequently used. The relative frequency or absence of terms across corpora is used as a proxy to assess the salience/familiarity of a word to all speakers of a variety. For comparison, we looked for the items in the *Corpus of Contemporary American English* (Davies, 2008), the largest linguistic corpus ever created, with over a billion words. *Bashful* occurs there 516 times, verifying its existence in American English, while underlining its low salience with a normalized frequency of about 0.5 per million words. Terms such as *hot-tempered* are likely uncommon because synonymous expressions such as "He's got a hot temper" are used more often.

Study 2.2: Unfamiliar Terms and Exploratory Factor Analysis in the ALLS

Using ALLS data, we evaluated items based on the endorsement of the response option "I don't know this word" to identify those that are unfamiliar in these contexts, and we conducted exploratory factor analysis (EFA) to evaluate dimensionality among the items.

Study 2.2: Method

Participants. A total of 3,048 mostly18-year-old participants, speakers of over 50 different languages, were recruited in 2022 in three countries, Namibia, Kenya, and South Africa, for a 5-year longitudinal study of personality development, mental health, and emerging adulthood. Of this sample, 3,027 participants filled out the CCB2I. Sample characteristics at Wave 1 are shown in Table 3. They include nonoccupational items to assess socioeconomic status from the Family Affluence Scale II (Boyce et al., 2006; Currie et al., 2008), and it also included details on housing situation, intended to give readers a sense for participants' context. More details about the sample and procedures are in Thalmayer et al. (2024).

Procedure. Research assistants (RAs) who are schoolteachers, sociology or psychology graduates or students, or experienced data collectors recruited participants in their local communities, including multiple regions of each country and both urban and rural areas. Participants provided written informed consent to join the study and then completed either an article or an electronic version of the survey, on a tablet provided by the RA, in the RA's presence. The language of the survey was English, the language of secondary schooling for all participants and thus the language they most likely read and write best, though most speak other languages in their family and social lives. Translation of instructions and measures (not including personality items) was available to RAs in some local languages to aid their support of participants. The 55 items from Study 1 were administered with the stem "I tend to be" and response options from 1 = *very inaccurate* to 6 = *very accurate*, plus the option, "I do not know this word." Additional information regarding these data is presented in Study 3.

Analyses. Participants were excluded from analysis if they failed three or four of four attention checks in the full survey or if they consistently chose the same answer option after an attention check, a common misunderstanding ($n_{\text{Kenya}} = 174$, $n_{\text{Namibia}} = 265$, $n_{\text{South Africa}} = 92$). In total, 531 participants were excluded according to these criteria, leading to an analysis sample of 2,496. The percent of "I do not know this word" responses was calculated for each item by country and as a total for the full data set. EFA with oblimin rotation was used to examine factor solutions in both raw and ipsatized data (i.e., to remove each person's mean response from each item response, to adjust for response styles like yea or nay saying), by country and for the combined data set (factor correlation in raw data = −.25, ipsatized data = .50). Of interest were items with a primary loading on the unintended factor, a low primary loading on the intended factor, and/or those with a low ratio of primary (intended) to secondary (unintended) loadings.

Study 2.2: Results and Discussion

"I do not know this word" responses accounted for about 7% of all responses. Average percentages for each term across the combined data are listed in ascending order in the second column of Table 2. The terms with the highest percentages (*benevolent*, *bashful*, *pessimistic*, *timid*, *industrious*, *courteous*) tended to be a bit longer than those with the lowest (e.g., *honest*, *brave*, *kind*). Results by country are reported in Supplemental Table S2. The use of this answer option varied between the contexts, from 13% in Namibia, to 5% in Kenya and 4% in South Africa. These differences do not alter the rank order of results, which was our primary interest. We suspect variation had less to do with vocabulary knowledge than with RA training. At every site, participants expressed reluctance to use this option and asked RAs to define the words. It became clear that the wording was too direct, especially as many participants completed the survey at school in the presence of teachers, which may have made them feel they were failing a test. Subsequent discussion with RAs and collaborators led us to adopt "unknown word" as more face-saving phrasing for future surveys.

Detailed EFA results are reported in Supplemental Table S3. The key results of factor loadings on the intended factor in combined ipsatized data and loading ratios are reported in the third and fourth

Table 2

Summary of Results for Study 2: Problematic Items Identified Across Methods and Samples

Ten corpora		Africa Long Life Study (ALLS)						SAPA project online data							
Number missing		% Rated unknown (*N*)		Loading intended factor		EFA ratio		DTS average *p*		Loading intended factor		EFA ratio		Test–retest *r*	
Bashful	10	**benevolent**	40	**Bashful**[a]	−.06	**Cowardly**[a]	1.02	Sad	.05	Naive	−.08	Hardworking[a]	1.20	**Humane**	.35
Good-natured	10	**bashful**	29	**Benevolent**	.06	Weak	1.04	Timid	.05	**Benevolent**	.25	Diligent	1.32	**Courteous**	.46
Hotheaded	10	pessimistic	27	**Industrious**[a]	.12	**Benevolent**	1.16	Unhappy	.06	Diligent	.27	Pessimistic	1.52	Kind	.52
Hot-tempered	9	timid	27	Modest	.13	Pessimistic	1.21	Friendly	.06	Gossipy	−.29	Daring	1.65	Faithful	.53
A loner	8	**industrious**	26	**Courteous**	.25	Unhappy	1.45	**Hotheaded**	.06	Hardworking	.30	Friendly	1.78	**Benevolent**	.53
Kindhearted	8	**Courteous**	21	Pessimistic	−.26	**Naive**	1.53	**Cowardly**	.06	**Industrious**	.31	Unhappy	1.81	Modest	.55
Gossipy	7	Dynamic	20	**Naive**	−.26	Clever	1.53	Silent	.06	Clever	.32	Modest	1.84	Good-natured	.56
Self-confident	6	**Hotheaded**	17	Diligent	.29	Determined	1.58	Aggressive	.06	Faithful	.32	**Bashful**	2.08	Peaceful	.57
Disobedient	6	Modest	17	Clever	.30	Timid	1.6	Peaceful	.06	**Bashful**	−.33	Cheerful	2.15	Diligent	.57
Industrious	5	Diligent	15	Dynamic	.30	Modest	1.84	Bad	.06	Modest	.34	Sad	2.20	Respectful	.58
Inhumane	4	**Naive**	14	**Cowardly**[a]	−.32	**Industrious**[a]	1.99	Brave	.07	Calm	.35	Happy	2.24	Disobedient	.59
Cowardly	3	Compassionate	11	Determined	.33	Happy	2.01	A loner	.07	Friendly	.36	Bold	2.33	Friendly	.59
Compassionate	3	Daring	11	Compassionate	.33	Diligent	2.2	Arrogant	.07	Honest	.37	Obedient	2.47	**Industrious**	.60
Courteous	3	Bold	10	Weak	−.34	Sad	2.28	Weak	.07	Generous	.40	Determined	2.70	Bad	.60
Humane	3	Loner	10	Timid	−.34	Aggressive	2.38	Quiet	.07	Obedient	.42	Gentle	3.22	Bold	.61
Courageous	2	**Humane**	10	Happy	.39	Daring	2.67	Gossipy	.07	Determined	.44	**Industrious**[a]	3.26	Generous	.62
Hardworking	2	Inhumane	10	Cheerful	.41	Hardworking	2.99	Shy	.07	Pessimistic	−.46	Honest	3.42	Timid	.63
Daring	1	Hot-tempered	7	Daring	.41	**Courteous**	3.36	Modest	.08	**Humane**	.46	Gossipy	3.43	Determined	.64
Pessimistic	1	Arrogant	7	**Humane**	.41	Quiet	3.36	**Industrious**	.08	Dynamic	.47	Calm	4.15	Shy	.64
Timid	1	Disobedient	7	Unhappy	−.42	Bold	3.4	Bold	.08	Peaceful	.48	A loner	4.26	Polite	.64
Benevolent	1	Obedient	7	Calm	.42	Calm	3.45	Disobedient	.09	Cheerful	.49	Hot-tempered	4.38	Courageous	.65
Diligent	1	**Cowardly**	6	Bold	.44	Hot-tempered	3.69	**Bashful**	.09	Inhumane	−.49	Weak	4.51	Weak	.65
Obedient	1	Courageous	5	Obedient	.44	Silent	3.84	Gentle	.09	**Courteous**	.49	Bad	4.52	Compassionate	.65
Polite	1	Good-natured	5	Hardworking	.45	**Hotheaded**	4.09	Inhumane	.09	A loner	−.50	Naive[a]	4.69	Calm	.66
Arrogant	1	Aggressive	4	Hot-tempered	−.46	Gossipy	4.46	Naive	.09	Daring	.51	**Cowardly**	5.16	Unhappy	.66
Active	0	Generous	4	Aggressive	−.46	**Bashful**[a]	4.5	Calm	.09	Compassionate	.51	Quiet	5.17	Dynamic	.67
Aggressive	0	Determined	4	Brave	.47	Cheerful	4.92	**Benevolent**	.09	Happy	.52	**Hotheaded**	5.24	Gentle	.67
Bad	0	Lively	4	Active	.47	Obedient	5.42	Courageous	.09	Active	.52	Polite	5.28	Obedient	.67
Bold	0	Gossipy	3	Sad	−.48	Polite	6.75	Obedient	.09	Disobedient	−.54	Brave	5.74	Energetic	.69
Brave	0	Energetic	3	Gentle	.48	Courageous	7.34	Happy	.09	Silent	−.56	Inhumane	6.08	Honest	.70
Calm	0	Cheerful	3	**Kindhearted**	.48	Kind	7.71	Hot-tempered	.09	Quiet	−.57	Dynamic	8.29	**Kindhearted**	.70
Cheerful	0	**Kindhearted**	2	Peaceful	.48	Shy	7.83	**Self-confident**	.10	Gentle	.57	Courageous	8.93	Inhumane	.70
Clever	0	Gentle	2	Lively	.49	Compassionate	7.91	Lively	.10	Respectful	.57	Confident	9.12	Happy	.70
Confident	0	Polite	2	Gossipy	−.49	Active	7.96	Confident	.10	Unhappy	−.58	Active	9.55	Lively	.70
Determined	0	Bad	2	**Hotheaded**	−.50	Energetic	9.15	Daring	.10	Good-natured	.58	Shy	9.64	Cheerful	.71
Dynamic	0	Unhappy	1	Courageous	.51	Loner	9.19	Determined	.10	Sad	−.59	**Self-confident**	10.21	Daring	.72
Energetic	0	Calm	1	Quiet	−.51	Respectful	9.66	Pessimistic	.11	Lively	.59	**Kindhearted**	11.59	Naive	.72
Faithful	0	Clever	1	Good-natured	.51	**Self-confident**	10.64	Active	.11	Polite	.59	Aggressive	11.72	Silent	.72
Friendly	0	**Self-confident**	1	Honest	.51	Lively	11.78	Energetic	.12	Energetic	.60	Generous	12.17	**Cowardly**	.72
Generous	0	Weak	1	Loner	−.52	Bad	15.5	Faithful	.12	Hot-tempered	−.60	Good-natured	12.25	Quiet	.73
Gentle	0	Quiet	1	Generous	.52	Faithful	16.61	Good-natured	.13	**Kindhearted**	.60	Respectful	12.42	Pessimistic	.73
Happy	0	Active	1	Polite	.52	Disobedient	17.82	Dynamic	.13	Arrogant	−.60	Kind	12.93	Active	.73
Honest	0	Friendly	1	**Self-confident**	.53	Confident	20.67	Cheerful	.13	Bad	−.61	**Humane**	13.04	Clever	.74
Kind	0	Shy	1	Energetic	.53	Brave	23.02	Diligent	.13	**Hotheaded**	−.61	Disobedient	15.16	**Self-confident**	.74
Lively	0	Faithful	1	Friendly	.53	Generous	23.39	**Kindhearted**	.14	**Cowardly**	−.62	Silent	15.51	**Bashful**	.74
Modest	0	Silent	1	Faithful	.53	Gentle	40.46	Polite	.14	Weak	−.63	**Benevolent**	15.57	Brave	.74

(*table continues*)

Table 2 (*continued*)

Ten corpora		Africa Long Life Study (ALLS)					SAPA project online data					
		% Rated unknown	Loading intended				DTS average	Loading intended		EFA		Test–retest
Number missing		(*N*)	factor		EFA ratio		*p*	factor		ratio		*r*
Naive	0	Happy	1	Shy	−.55	Dynamic	47.39	**Courteous**	.14	Brave	.63	Clever	16.99	Hot-tempered	.75
Peaceful	0	Confident	1	Silent	−.56	Honest	52.05	Generous	.15	Aggressive	−.63	Lively	24.66	Confident	.75
Quiet	0	Sad	1	Respectful	.58	Arrogant	54.04	Clever	.15	Bold	.64	Timid	31.13	**Hotheaded**	.76
Respectful	0	Respectful	1	Confident	.59	Friendly	74.76	Hardworking	.16	Kind	.64	Energetic	31.52	Aggressive	.77
Sad	0	Honest	1	Arrogant	−.61	**Kindhearted**	79.11	Respectful	.16	Courageous	.65	Peaceful	46.58	Gossipy	.78
Shy	0	Hardworking	1	Inhumane	−.62	Peaceful	86.57	Kind	.17	Timid	−.66	Compassionate	52.20	Arrogant	.78
Silent	0	Kind	1	Kind	.64	Good-natured	108.13	**Humane**	.18	Shy	−.67	**Courteous**	61.68	A loner	.78
Unhappy	0	Brave	1	Disobedient	−.66	Inhumane	427.56	Compassionate	.19	**Self-confident**	.70	Arrogant	92.92	Sad	.78
Weak	0	Peaceful	1	Bad	−.68	**Humane**	1452.99	Honest	.21	Confident	.73	Faithful	121.79	Hardworking	.80

Note. In all columns, terms are in ascending order of most problematic. The 10 items ultimately chosen for removal are in bold. Study 2.1: Value shown is the number of corpora, of 10, from which the term was missing. Study 2.2 ALLS ($N = 2{,}496$): Value shown is percentage of participants who chose the option "I do not know this word" for the item, averaged across three samples (Namibia, Kenya, South Africa). The EFA values here and for Study 2.3 are first the loading on the intended factor and then the ratio between it and the unintended factor. Study 2.3 SAPA data ($N = 61{,}760$): For DTS, the value shown is average *p* value for tests of difference between each pair of 13 countries. Test–retest values are within session (see Study 2.3). SAPA = Synthetic Aperture Personality Assessment; EFA = exploratory factor analysis; DTS = distribution test statistics.

[a] Had their primary loading on the unintended factor.

columns of Table 2, shown from lowest to highest. Three items, *cowardly*, *industrious*, and *bashful*, had their primary loading on the unintended factor. Several others, for example, *weak*, *benevolent*, *pessimistic*, *modest*, *naïve*, and *unhappy*, had low ratios, indicating weaker specific validity for their intended factor.

Of the 10 least-known words in the ALLS, only three, *bashful*, *hot-headed*, and *industrious*, overlap with the 10 items least frequently found in Study 2.1 corpora. This may have to do with differences between young adults and the broader populations represented in the corpora, as well as with regional differences. *Good-natured*, which did not appear in any corpus outside the United States, was not especially unknown in ALLS samples. This may be because the term, while not frequently written, is morphologically transparent and easy to interpret.

The term *bashful* was brought to our attention early in the ALLS by schoolteacher–RAs who themselves did not know it. In a meeting of the four American coauthors, we realized that, although the word seemed familiar (perhaps due to an eponymous dwarf in Snow White's entourage?), none of us could define its semantic difference from *shy*. A dictionary informed us that it means holding back out of modesty rather than the social anxiety implied by shyness. While this distinction is interesting psychologically, it is not useful for a self-report scale if no one knows it.

Study 2.3: Exploratory Analyses Using Online Data From 13 English-Speaking Countries

In online survey data from 13 English-speaking countries in five world regions, three types of exploratory analyses were used to flag items for possible removal.

Study 2.3: Method

Participants. The total sample included 103,574 respondents from 234 nations who visited the data collection platform at https://sapa-project.org (SAPA). These respondents self-reported fluency in English and rated themselves on an average of 19.7 of the 55 descriptors ($Mdn = 16$; $SD = 14.6$; minimum = 1; maximum = 55). From this sample, respondents were filtered to include only those from nations with 200 or more respondents and which use English as an official or de facto language (Central Intelligence Agency, 2021). In total, this subset included 61,760 respondents from 13 nations who ranged in age from 18 to 90 years ($M = 25.8$, $SD = 9.0$; $Mdn = 23$). General demographic information is provided by nation in Table 4; additional information about the sample(s) is provided in the SAPA additional online materials available at the OSF at https://osf.io/89x2q/.

Procedure. Respondents completed a cross-sectional, self-report survey in exchange for customized feedback about their personality traits. Each respondent was administered a subset of items randomly drawn from a large pool that included CCB2I items, formatted with the stem and response options as described in Study 2.2. This planned-missingness design results in data that are "missing completely at random" (Condon et al., 2017; Revelle et al., 2017). In addition to the sample size based on the number of respondents, an important consideration of this procedure relates to the number of pairwise administrations for each item pair. In the participants used in the present study, the *M* of pairwise administrations was 11,529, with a minimum of 10,034 ($SD = 339.8$). This allowed for the collection of

Table 3
Africa Long Life Study Sample Characteristics at Wave 1

Characteristic	Kenya (n = 739)		Namibia (n = 891)		South Africa (n = 836)		Total (N = 2,466)	
	n	%	n	%	n	%	n	%
Age = 18	623	84.4	742	83.7	608	73.0	1,973	80.3
19+	115	15.6	145	16.3	225	27.0	485	19.7
Gender = female	368	49.8	489	54.9	566	67.9	1,423	57.8
Other					4	0.5	4	0.2
Relationship = single	545	73.7	649	73.0	644	77.0	1,838	74.6
Parent status = no children	726	98.2	817	91.9	799	95.9	2,342	95.2
One child	10	1.4	67	7.5	30	3.6	107	4.3
Two children	1	0.1	4	0.4	2	0.2	7	0.3
>Two children	2	0.3	1	0.1	2	0.2	5	0.2
Full-time student	625	84.6	763	85.7	694	83.2	2,082	84.5
Education of female caregiver								
Grade 10 or less	309	41.8	478	53.9	207	25.1	994	40.6
Grade 11 or 12 secondary	162	21.9	187	21.1	269	32.6	618	25.2
Vocational training	65	8.8	58	6.5	112	13.6	235	9.6
Some university or diploma	94	12.7	55	6.2	173	21.0	322	13.1
University bachelor's degree	60	8.1	32	3.6			92	3.8
Masters/postgrad degree	26	3.5	23	2.6	63	7.6	112	4.6
No female caregiver	23	3.1	54	6.1			77	3.1
Housing situation								
House/apartment owned by family	333	45.1	524	59.6	592	71.5	1,449	59.2
House/apartment rented by family	185	25.0	107	12.2	125	15.1	417	17.0
Informal/nonpermanent structure	182	24.6	163	18.5	76	9.2	421	17.2
Live in someone else's home	39	5.3	85	9.7	35	4.2	159	6.5
Have electricity	535	72.4	662	75.2	816	97.8	2,013	82.1
Have a smartphone[a]	502	53.6	523	44.7	526	56	1,551	50.9
Have own bedroom	395	53.5	405	46.0	475	56.9	1,275	51.9
Have computer in household	252	34.1	305	34.7	547	65.6	1,104	45.0
Have car in household	185	25.0	278	31.6	494	59.2	957	39.0

[a] Determined by checking phone numbers given for a WhatsApp account, which is ubiquitously used in these contexts, but only accessible with a smartphone.

data on several other self-report scales that are used and described in Study 4.

Analyses. DTS (also Kolmogorov–Smirnov, Cramer-von Mises, Anderson-Darling, and Wasserstein distance tests) were conducted to evaluate the probability that responses from different countries are samples drawn from a common population of respondents. First, observed response data is transformed into empirical cumulative distribution functions. Then, we compare the response distributions for each item across pairs of countries using five nonparametric tests using the distCompare function for scaling the analyses and the twosamples package in R (Dowd, 2020, 2022). Significant differences between nations in the response pattern for an item suggest that one or more characteristics of the item cause it to function differently across the two groups. The absence of significant differences suggests that the item functions similarly. SAPA Supplemental Materials include more details, an example, and the differences between the five tests.

EFA was used to evaluate dimensionality and other structural properties. Though a two-factor structure was anticipated, we sought to replicate the expected dimensionality with EFA rather than confirmatory factor analysis (CFA) consistent with recommendations (e.g., Costello & Osborne, 2005). EFA with oblimin rotation was used to examine factor solutions with two to four factors in both raw and ipsatized data. Finally, within-session test–retest reliability analyses were used to identify items with extreme consistency, whether high or low.

Study 2.3: Results

DTS results are summarized by listing mean p values for each test by item, in the fifth column of Table 2, sorted by ascending p value. The correlations among p values across test statistics were very high (r = .87–.99). Results for all five tests by item and pair of nations are in Supplemental Materials. The items with the largest differences across nation pairs were *sad*, *timid*, *unhappy*, *friendly*, and *hot-headed*. The most consistent response patterns across contexts were for *honest*, *compassionate*, *humane*, *kind*, and *respectful*. Agency terms, especially related to negative emotions, appear to be more variable across contexts than Communion/SSR terms. Aggregation of the p values for all items across nation pairs indicated relatively similar response patterns among Australia, Canada, Great Britain, and the United States. By contrast, Kenya, the Philippines, and, to a lesser extent, India, differed the most from one another.

The sixth column of Table 2 includes results from EFA with ipsatized data, specifically the factor loading on the intended factor and the ratio of primary to secondary loadings. Further results from EFA using raw and ipsatized data are in SAPA additional online materials. Three items intended for the Communion/SSR scale (*industrious*, *hardworking*, and *naïve*) had their primary loading on Agency/Dynamism. Of these, *hardworking* was especially interstitial (i.e., located between the two dimensions), with very similar primary and secondary loadings. *Industrious* and *naïve*, on the other hand, were

Table 4
SAPA Project Sample Characteristics

Country	*N*	Age		%			
		Range	*M* (*SD*)	Female	Nonbinary	Student	Degree
United States	13,313	18–90	28.9 (12.3)	65.5	0.6	22.5	14.5
Australia	633	18–78	31.8 (12.7)	68.7	0.3	13.7	18.8
Canada	1,042	18–79	30.8 (13.1)	62.6	0.3	20.7	18.3
India	17,903	18–79	24.0 (6.1)	72.0	0.1	38.7	19.2
Jamaica	1,514	18–70	25.9 (8.5)	81.3	0.1	21.2	14.0
Kenya	5,913	18–79	25.1 (7.2)	68.0	0.1	33.9	22.8
Namibia	744	18–69	25.7 (7.5)	75.7	0.0	26.6	16.7
Nigeria	3,389	18–69	24.4 (6.6)	67.2	0.2	40.1	21.5
Pakistan	3,587	18–76	24.1 (6.6)	71.0	0.1	42.0	19.3
Philippines	8,320	18–79	23.5 (6.2)	81.9	1.1	37.9	24.3
South Africa	2,104	18–79	28.6 (10.6)	77.5	0.3	25.1	16.0
Trinidad and Tobago	1,643	18–90	27.8 (9.5)	77.4	0.1	16.9	14.3
Great Britain	1,655	18–86	31.0 (12.9)	63.5	0.3	17.3	25.3

Note. SAPA = Synthetic Aperture Personality Assessment; Student = identified as currently a full-time student (type of school not specified); Degree = has a college degree.

more dominantly aligned with Communion/SSR, the unintended scale. In general, the most clearly oriented items in the two-factor solution—those with a high ratio of primary:secondary loadings (>3)—were Communion/SSR items. Most items with a low (<3) primary: secondary ratio were Agency/Dynamism items (10 out of 13 in ipsatized data).

Within-session test-retest reliability is shown on the far right of Table 2, from lowest to highest scores. Most items had good values, at or above $r = .60$. The worst performers were *humane*, *courteous*, *kind*, *faithful*, *benevolent*, and *modest*.

Study 2.3: Discussion

Three types of analysis in online data from 13 countries were used to identify items that did not work well across English-speaking contexts. The results of the tests were variable. DTS analyses highlighted variability in how emotion-laden Agency/Dynamism terms were responded to, EFA highlighted that conscientiousness-related terms combining approach–behavior with social responsibility are interstitial in the CCB2, and test–retest analyses highlighted the lower reliability of character-relevant Communion items, some of which may be unfamiliar.

Study 2: Integration and Discussion

Sources of data from many nations were compiled to identify the most and least consistently functioning of the 55 CCB2I items generated from emic (localized, bottom–up) lexical studies. Given their different emphases, it is not surprising that these sources of information identified different items as more and less problematic. The variability of these results underlines the merit in retaining a relatively large number of items until translations into diverse languages can further inform removal/retention decisions.

At this stage, we found the accumulated evidence to support the removal of 10 items. *Bashful* and *benevolent* were removed due to their unfamiliarity in the ALLS: the former also due to its absence in World English corpora and the latter also due to its poor retest reliability. The presence of both *bashful* and *shy* in Study 1 results may indicate that their distinction is more salient in other languages, leading to translations matching both English terms. Thus, future translations should consider including both kinds of content, if available in locally familiar vocabulary. (Luckily, we do maintain the term *modest*.) Three other problematic items—*hot-headed*, *kind-hearted*, and *self-confident*—were easy to drop because simpler, less metaphoric synonyms were among retained items: *hot-tempered*, *kind*, and *confident*. *Humane*, which had the worst retest reliability, is better represented by its more familiar antonym, *inhumane*. Similarly, *courteous*, which had the second worst retest reliability and was highly unknown in the ALLS, is better represented by its close synonym *polite*. *Industrious* was the most interstitial item, and its close synonyms, *hardworking* and *diligent*, performed better on all metrics. *Cowardly* was also interstitial and had the highest loading on the unintended factor; it is better represented by close antonyms *courageous* and *brave*. *Naïve* was hard to categorize in Study 1, unfamiliar in the ALLS, and unassociated with its intended factor in SAPA data. These 10 terms are presented bold in Table 2.

While *gossipy* was not especially problematic by most metrics, it was missing from most corpora, and RAs and coauthors remarked on how seldom the term is used in this form. A post hoc assessment of *gossip* in the 10 corpora found it present in eight as a verb (including the nominalization "gossiping"). Though it was present in all 10 corpora as a noun, evidence does not suggest its consistent use as a person descriptor. Thus, the term was changed to verb form (e.g., "I tend to gossip") for future data collection. Several other items teeter on the edge of acceptability, but as stated from the outset, we chose to maintain a larger rather than smaller pool to allow for future refinements after translations and use in more contexts and thus only excluded 10 items from the next two studies.

Study 3: The CCB2I in the ALLS

The goal of Study 3 was to assess the refined 45-item CCB2I, displayed in Table 5, in the ALLS. A 36-item version with balanced keying was also assessed. Psychometric properties and

Table 5
The 45-Item Cross-Cultural Big Two Inventory

Communion/Social Self-Regulation			Agency/Dynamism		
Component	Forward	Reversed	Component	Forward	Reversed
Warmth	Compassionate[a]		Competence	Clever	Weak
	Friendly[a]		Confidence	Confident	Timid
	Generous		Fearlessness	Brave	Cowardly[b]
	Good-natured[a]			Bold	
	Kind			Courageous[a]	
Morality	Faithful[a]	Bad		Daring[a]	
	Honest	Inhumane		Determined	
Respect	Modest[a]	Arrogant	Positive mood	Cheerful	Pessimistic
	Obedient[a]	Disobedient		Happy	Sad
	Polite	Gossipy			Unhappy
	Respectful		Sociability		A loner
Industriousness	Diligent[a]				Shy
	Hardworking		Surgency	Active	Quiet
Even temper	Calm[a]	Aggressive		Dynamic[a]	Silent
	Gentle	Hot-tempered		Energetic	
	Peaceful[a]			Lively	

Note. The stem "I tend to be" is used with all items. Scales are scored as the mean response of items answered, after reverse scoring relevant items. In the Africa Long Life Study, response options ranged from 1 = *very inaccurate* to 6 = *very accurate*, with the additional option "unknown word." The 45 items include 23 Communion/Social Self-Regulation and 22 Agency/Dynamism terms.
[a] To create a 34-item version with balanced keying, these 12 items were removed. [b] *Cowardly* was added to reverse-keyed Agency/Dynamism items, for a total of 14 Social Self-Regulation and 20 Agency/Dynamism items.

associations of items and scales with mental and physical health and religious engagement are tested in the samples of 18-year-olds in Namibia, Kenya, and South Africa described above for Study 2.2.

Study 3: Method

In addition to CCB2I items, the ALLS Wave 1 survey included the measures of mental and physical health and religious engagement, the demographic and socioeconomic variables reported in Table 3, and two open-ended items about participants' interests in and questions about the study that are not reported here.

Materials

The International Mental Health Assessment (IMHA; 60 items; Thalmayer et al., 2023), assesses psychological problems scored in 10 categories, which are grouped into three overarching spectra. Six subscales and two spectra are used and reported here: Internalizing (Depression, Anxiety, Posttraumatic Stress, Attention Deficit/Hyperactivity) and Externalizing (Substance Abuse, Anger, Attention Deficit/Hyperactivity). The IMHA facilitates cross-cultural comparison at the item and scale levels by assessing symptom frequency from 0 = *not in the last month/never* to 6 = *every day*.

The The Duke Religion Index (five items; Koenig et al., 1997) assesses religiousness and spirituality, distinguishing between intrinsic religiosity and religious activity, including one item for organizational and one for nonorganizational activity. The latter are often combined, which is done here as previously validated for use in Namibia (Toscanelli et al., 2022).

The one-item General Self-Reported Health (DeSalvo et al., 2005) allows participants to rate their health on a 5-point scale ranging from *poor* to *excellent*.

Hypotheses

Based on associations established largely in Western contexts, we anticipated that higher scores on Communion/SSR would associate with identifying as female, better physical and mental health especially in terms of lower Externalizing spectrum subscale scores, and higher religious engagement and that higher scores on Agency/Dynamism would associate with identifying as male and lower internalizing scores but potentially higher substance use.

Analyses

Scale scores for the CCB2I are the mean response of items answered. Psychometric properties, including mean interitem and item–total correlations, Cronbach's α are used to assess the inventory's reliability in these contexts and age group. The omega hierarchical (ω_h) is used to assess general factor saturation, that is, how much of the content is explained by the primary dimension and thus the degree to which the group of items assesses a single underlying construct and is unidimensional (Revelle & Condon, 2019). Zero-order and partial correlations (controlling for the overlap of the CCB2I scales) for age, gender, mental health, physical health, and religious engagement are used to test the hypotheses described above. In addition to the full 45-item version, analyses were conducted with a 34-item balanced-key (BK) version, shown in Table 5, and a 45-item version with acquiescence responding removed. The latter was achieved by identifying five pairs of near antonyms with minimum .21 magnitude negative correlation (*obedient/disobedient*, *happy/sad*, *kind/inhumane*, *good-natured/bad*, *cowardly/courageous*). Item responses were adjusted at the individual level using the average response and standard deviation on these 10 items.

Three levels of measurement invariance are explored using lavaan (v. 0.6-15; Rosseel, 2012), following standard multigroup CFA

procedures (Byrne & Van De Vijver, 2010; Fischer & Karl, 2019; Gregorich, 2006; McDonald & Ho, 2002): (a) configural equivalence (parameters free to vary for each group), (b) full or partial metric invariance (constraining factor loadings to be equal between groups and individual item testing where invariance does not hold), and (c) scalar invariance (constraining item intercepts to equality across groups). For configural CFA fit, we compared results not to standard benchmarks in statistical literature but to fit for similar multidimensional personality inventories as recommended by Hopwood and Donnellan (2010), specifically to those they report for eight inventories constructed and tested in North America. Two had inadmissible results, while the remaining six had the following: Tucker–Lewis index (TLI) .52–.70, comparative fit index (CFI) .61–.79, root-mean-square error of approximation (RMSEA) .09–.13. If it is difficult to achieve good fit of multidimensional models in one population, it will be even more challenging to find it in models tested across diverse populations. Thus, comparisons are made to these fit indices as reasonably high, "domain-specific" benchmarks. Another point of comparison is a 14-item two-factor model tested by Thalmayer and Saucier (2014) in eight nation samples: TLI .43–.78, CFI .53–.82, RMSEA .06–.10.

Study 3: Results

Psychometric properties are reported in Table 6 and in more detail in Supplemental Table S4. Reliability in terms of Cronbach α was good for the CCB2I scales in all contexts, with the only value under .80 for Agency/Dynamism in Namibia and stronger values for the initial 45-item version compared to the two alternatives. The ω_h values indicated evidence for moderate unidimensionality in the Communion/SSR scale, but those for Agency/Dynamism are below the values typically seen in scales with high unidimensionality. This result is unsurprising given the breadth of content covered with a two-dimensional measure of content with very broad scope. It should be expected that two-factor scales with adequate breadth to cover the intended domain have relatively low unidimensionality, consistent with other low dimensional, broad-bandwidth measures of personality (Revelle & Condon, 2019). The IMHA and The Duke Religion Index scales had fewer items per scale and lower reliability, which was generally similar across samples. The ω_h values were acceptable for most scales, with the exception of the three-item Intrinsic Religiosity scale.

The big two scales were correlated .53 with each other in the original version, .51 for the balanced key version, and .16 for the acquiescence-adjusted version. Thus, associations between the CCB2I scales and gender, physical and mental health, and religious engagement collapsed across countries, are most interpretable in the acquiescence-adjusted version (Figure 1) and are highly similar to partial correlations shown in Supplemental Figure S1. Zero-order correlations in the original and BK version, which are highly similar to each other but differ substantially from those shown in Figure 1, are also in Supplemental Figure S1. In acquiescence-adjusted correlations, women scored slightly higher on Communion/SSR and men on Agency/Dynamism. Associations with Communion/SSR were, as expected, in terms of lower Externalizing scale scores, but they also trended toward very slightly lower self-reported physical health and higher Internalizing tendencies. Associations with Agency/Dynamism were, as expected, in terms of better self-reported health and lower Internalizing scale scores but slightly higher risk of substance use and anger. Intrinsic religiosity was slightly higher, as anticipated for Communion/SSR, but the association with religious activity was stronger for Agency/Dynamism. Examination of the components of organizational and nonorganizational religious activities indicated little difference in associations. Zero-order and partial correlation coefficients by country are shown in Supplemental Tables S5 and S6.

Table 6

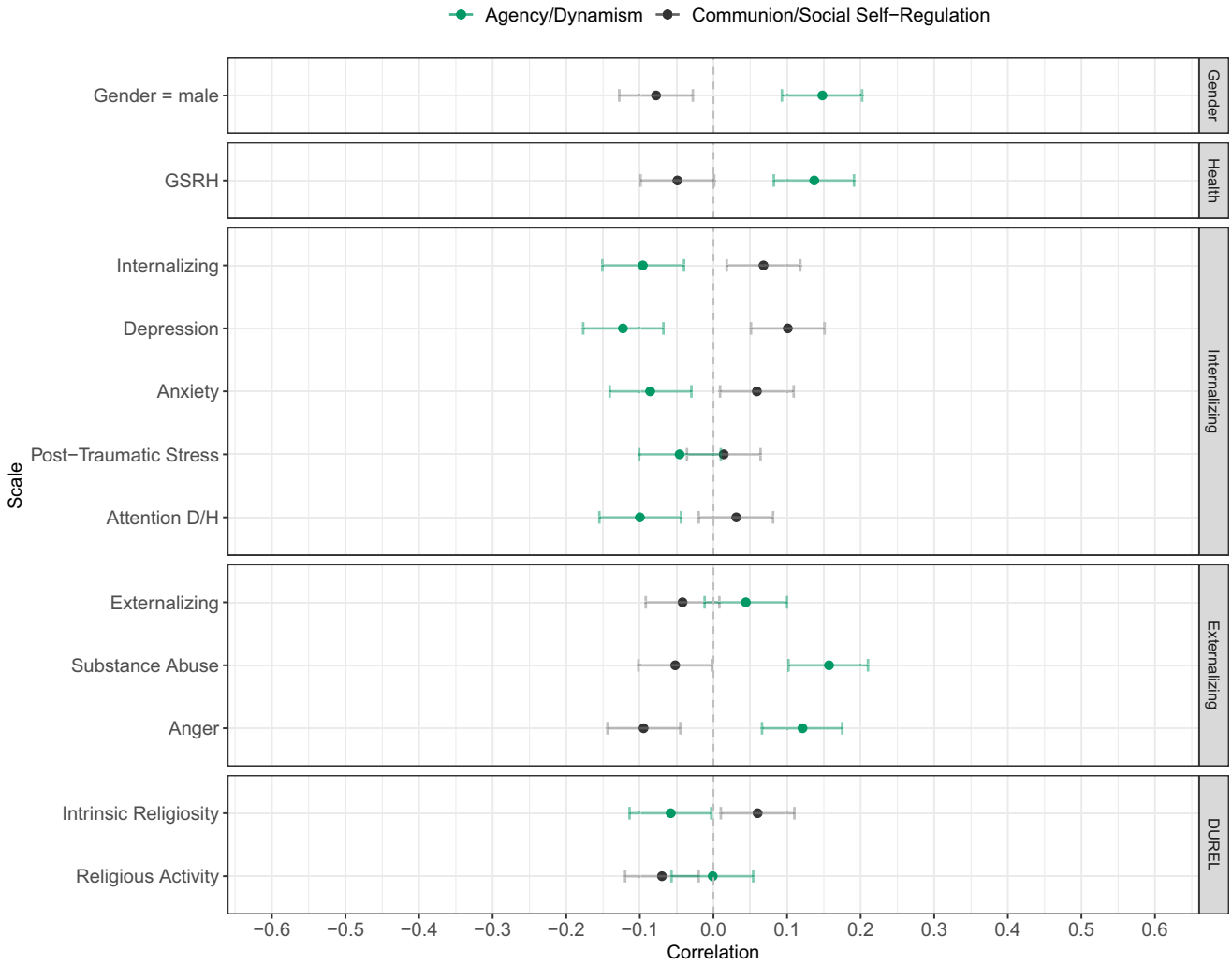
Psychometric Properties of the Cross-Cultural Big Two Inventory and Other Scales in Africa Long Life Study Wave 1 Data

	M			*SD*			α			ω_h		
Scale	K	N	SA	K	N	SA	K	N	SA	K	N	SA
45-item Communion/SSR	4.33	4.13	4.33	1.10	1.49	1.30	.90	.91	.90	.73	.82	.62
45-item Agency/Dynamism	4.15	4.15	4.27	1.34	1.54	1.47	.87	.78	.85	.41	.24	.35
ACQ Communion/SSR	0.32	0.24	0.26	0.56	0.82	0.70	.86	.86	.85	.67	.63	.77
ACQ Agency/Dynamism	0.27	0.25	0.24	0.68	0.86	0.80	.80	.65	.82	.25	.25	.45
BK Communion/SSR	3.66	3.71	3.82	1.17	1.52	1.35	.84	.84	.83	.64	.45	.50
BK Agency/Dynamism	3.94	4.02	4.15	1.35	1.56	1.49	.86	.75	.84	.48	.27	.41
Self-Reported Health	2.36	2.23	2.21	2.36	2.23	2.21						
IMHA Internalizing	1.96	2.19	2.43	0.92	1.01	1.11	.86	.86	.85	.84	.85	.80
Depression	2.03	2.30	2.58	1.36	1.51	1.63	.80	.75	.79	.72	.70	.70
Anxiety	1.99	2.27	2.47	1.36	1.55	1.63	.76	.75	.78	.69	.62	.66
Posttraumatic Stress	1.92	2.09	2.20	1.29	1.43	1.50	.72	.74	.70	.67	.70	.65
Attention Deficit/Hyperactivity	1.83	1.97	2.27	1.30	1.41	1.59	.75	.78	.71	.71	.76	.63
IMHA Externalizing	1.61	1.83	1.76	0.75	0.96	0.85	.67	.74	.58	.52	.53	.50
Substance Abuse	1.17	1.43	1.34	0.66	1.07	0.91	.85	.86	.80	.83	.85	.59
Anger	1.80	2.05	1.60	1.20	1.49	1.14	.75	.76	.74	.60	.63	.64
Intrinsic Religiosity	3.83	3.37	3.76	1.37	1.45	1.35	.64	.60	.73	.60	.58	.06
Religious Activity[a]	4.63	3.60	3.65	1.14	1.39	1.37	.49	.42	.48			

Note. $N_{\text{Kenya}} = 753$; $N_{\text{Namibia}} = 897$; $N_{\text{South Africa}} = 846$. Attention Deficit/Hyperactivity was included on both Internalizing and Externalizing spectrum scales. α = α on standardized items; ω_h = hierarchical omega; K = Kenya; N = Namibia; SA = South Africa; SSR = Social Self-Regulation; ACQ = 45-item version of the scales adjusted for acquiescence responding; BK = balanced keyed, 36-item version; IMHA = International Mental Health Assessment.
[a] Religious Activity has two items, and reliability cannot be calculated, so correlations between the items are shown under α.

Figure 1

Zero-Order Correlations Between Acquiescence-Adjusted CCB2I Scales and External Scales, Combined ALLS Samples



Note. CCB2I = Cross-Cultural Big Two Inventory; ALLS = Africa Long Life Study; GSRH = general self-reported health; Attention D/H = Attention Deficit/Hyperactivity. See the online article for the color version of this figure.

The details of CFA for all three versions of the measure and of measurement invariance tests in the original and acquiescence-adjusted 45-item versions are reported in Table 7. CFA fit for the initial and acquiescence-adjusted versions in each of the three samples was highly similar and above the benchmarks from Hopwood and Donnellan (2010) in terms of TLI and RMSEA and on par for CFI and for those reported by Thalmayer and Saucier (2014). CFI was expected to be low here because of the low saturation and low loadings due to moderate interitem correlations in multidimensional personality inventories. The fit was consistently poorer for the BK version. Fit indices for tests of measurement invariance across the three countries imply justification for accepting metric but not scalar invariance.

Study 3: Discussion

In Study 3, we assessed the measurement properties and associations of the 45-item CCB2I in the three samples of the ALLS. In all three samples, the CCB2I scales were more intercorrelated than the big two theory presupposes, limiting their differential prediction. This was successfully addressed using an acquiescence-adjusted version of the full scale, though not by an attempt to create a shorter, balance-key version. We found acceptable psychometric properties and metric-level invariance across groups for the original and the acquiescence-adjusted 45-item versions. Associations, most interpretable using acquiescence-adjusted scores, were as seen in Western research for mental health, with Communion/SSR predominantly predicting lower Externalizing spectrum scores and Agency/Dynamism predicting lower Internalizing scores. Associations with gender were very small but in predicted directions.

For physical health, there was a small positive correlation with Agency/Dynamism and not as predicted for Communion/SSR. This may be explained by the young age of the sample. At 18, personality tendencies related to health behaviors have had little chance to impact actual health. Choices such as alcohol use are restricted by

Table 7

Confirmatory Factor Analysis Fit in and Measurement Invariance Across Three Africa Long Life Study Samples for the 45-Item Cross-Cultural Big Two Inventory

Sample or test	χ^2	*df*	RMSEA	RMSEA 90% CI	SRMR	CFI	TLI	ΔCFI	ΔTLI
CFA fit by country and inventory									
Kenya	4135.57*	944	.067	[.065, .069]	.072	.730	.717		
Namibia	3896.94*	944	.059	[.057, .061]	.074	.783	.773		
South Africa	5280.28*	944	.074	[.072, .076]	.072	.688	.673		
BK Kenya	2809.24*	526	.077	[.074, .079]	.080	.701	.681		
BK Namibia	3115.94*	526	.075	[.072, .077]	.094	.713	.694		
BK South Africa	4265.32*	526	.092	[.090, .095]	.110	.602	.576		
ACQ Kenya	4120.00*	944	.068	[.065, .070]	.072	.727	.713		
ACQ Namibia	3881.67*	944	.059	[.057, .061]	.074	.783	.773		
ACQ South Africa	5229.42*	944	.074	[.072, .076]	.092	.690	.675		
Measurement invariance for original 45-item scale									
Configural invariance	15049.95*	2,832	.072	[.071, .073]	.08	.686	.670		
Communion/SSR	5174.92*	690	.088	[.086, .091]	.08	.778	.756		
Agency/Dynamism	5514.15*	627	.097	[.094, .099]	.09	.652	.616		
Metric invariance	15916.57*	2,918	.073	[.072, .074]	.10	.665	.660	−.02	−.01
Communion/SSR	5460.56*	734	.088	[.086, .090]	.09	.766	.758	−.01	0
Agency/Dynamism	6070.12*	669	.099	[.096, .101]	.12	.616	.602	−.04	−.01
Scalar invariance	18368.47*	3,004	.078	[.077, .080]	.11	.605	.609	−.06	−.05
Communion/SSR	6644.65*	778	.095	[.093, .097]	.10	.709	.716	−.06	−.04
Agency/Dynamism	7308.45*	711	.106	[.103, .108]	.12	.530	.542	−.09	−.06
Measurement invariance for ACQ scale									
Configural invariance	15015.63*	2,832	.073	[.071, .074]	.07	.603	.583		
Communion/SSR	3969.09*	690	.076	[.074, .079]	.06	.770	.747		
Agency/Dynamism	4895.18*	627	.091	[.089, .094]	.09	.599	.557		
Metric invariance	15717.47*	2,918	.073	[.072, .074]	.08	.583	.575	−.02	−.01
Communion/SSR	4227.50*	734	.076	[.074, .079]	.07	.755	.747	−.02	0
Agency/Dynamism	5342.88*	669	.092	[.090, .093]	.10	.561	.545	−.04	−.01
Scalar invariance	17271.49*	3,004	.076	[.075, .077]	.09	.535	.540	−.05	−.04
Communion/SSR	5072.06*	778	.082	[.080, .084]	.08	.699	.706	−.06	−.04
Agency/Dynamism	6022.19*	711	.096	[.093, .098]	.11	.501	.514	−.06	−.01

Note. $N_{\text{Kenya}} = 753$: $N_{\text{Namibia}} = 897$; $N_{\text{South Africa}} = 846$; $N_{\text{total}} = 2{,}496$. The first row at each level of invariance is for a model with both factors. Fit for each dimension is then shown separately. RMSEA = root-mean-square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual; CFI = comparative fit index; TLI = Tucker–Lewis index; BK = balanced keyed, 36-item version; ACQ = 45-item version with acquiescence responding removed; SSR = Social Self-Regulation.

* $p < .001$.

financial constraints and by living in hostels and with caregivers. It is also plausible that even in the long term, associations in these samples may be attenuated by the smaller range of health-relevant choices in these contexts, for example, in dietary options.

In terms of religious engagement, Communion/SSR was more associated with intrinsic religiosity and Agency/Dynamism with religious activity in the acquiescence-adjusted and partial correlations. In the zero-order correlations, all associations are positive and significant but without clear differentiation by trait or subscale. Gebauer et al. (2013) found that higher Agency/Dynamism predicted lower religious engagement in religious contexts, which they linked to the motivation of agentic individuals to differentiate from their ambient culture. This was not seen here, where Agency/Dynamism instead simply predicted the more extraverted side of religiosity. This difference could again be due to the age of the sample and their lack of social independence at age 18. However, we think it likely that this finding will not replicate in African contexts, where over 95% of the population belongs to a religion. With so few nonreligious individuals, personality differences are more likely to predict different kinds of religious engagement and activities rather than presence or absence. Although confidence and fearlessness could be helpful in differentiating from the ambient culture where there is a motivation to do so, we do not see that urge as intrinsic to Agency or implied by the content of the scale.

These findings must be considered in the context of limitations of the data and measures. First, the measures for external validity were created in the West. We chose measures optimized and tested for cross-cultural use, as noted above, and the ALLS study team gave careful attention to contextualization of item phrasing and to choosing topics of cross-culturally universal relevance (physical health) or African-specific relevance (religiosity). However, none of these measures were developed in or specific to the African context.

A second concern is the varying degree of English-language ability of ALLS participants, for most of whom English is a second language, used in schools and most written communications but not typically spoken at home or socially. We addressed this by giving participants the option to indicate a word as unfamiliar for CCB2I items but not for other surveys. It was also mitigated through the

vetting of wording by the ALLS team and by administration of surveys in the presence of local research assistants trained to understand the survey and with translations for their use to explain terms in several local languages. Given the 50 home languages of our sample, however, our research assistants were not able to speak all languages of their recruits. As for a written survey, there was no real alternative to English. Many of the languages of our participants are not written, and even where they are (typically as of the last century), participants are unlikely to have learned to read and write in them. Since the terms in the CCB2I were generated from a process including studies in three African languages, representing the three main language families of Africa and highly diverse societies (e.g., Thalmayer, Job, et al., 2021), the concepts are very likely to be transferable and to have similar valence across contexts. However, results must be interpreted knowing that the language and terms used in the CCB2I and other surveys are not words that participants grew up hearing in their familial and neighborhood contexts.

Another concern is the generally lower reliabilities and associations in Namibia compared to the two other countries. This could be due to the more rural and less globalized nature of many communities in Namibia, the second least densely populated country in the world, where many communities engage in aspects of traditional lifestyles, including subsistence farming. This population may be the most culturally distant of the three from the Western contexts that have shaped modern psychology (e.g., Henrich et al., 2010; Thalmayer, Toscanelli, et al., 2021).

A fifth notable limitation is the moderate measurement invariance. While CFA fit was better than domain-specific benchmarks, measurement invariance was not achieved at the scalar level, which prevents us from comparing means scores across samples. We knew from the outset that we were in a poor position to achieve good CFA properties due to the breadth and "lumpiness" of our scales and also that we would not want to drop items to shape our scales to improve fit, as this would narrow them and reduce content coverage at too early a stage, before translations and adaptation to other contexts.

There are two final limitations: the narrow age range of ALLS samples and the lack of other personality inventories included to assess comparative reliability and predictive power and thereby to disambiguate methodological challenges from potential limitations of the CCB2I. These are addressed by Study 4.

Study 4: Validation of the CCB2I in Multiple National Samples

The 45-item CCB2I is tested here in large samples from 13 diverse English-speaking countries to assess validity in a range of regions and contexts, as well as ages. Additionally, the model is compared to other personality measures, including two-factor models, and to outcome variables. The procedure and participants for this study are as described for Study 2.3.

Study 4: Method

Materials

In addition to the demographics and CCB2I items described in Study 2, participants responded to items from seven other personality scales using the same 6-point response format ("very inaccurate," "moderately inaccurate," "slightly inaccurate" … "very accurate"). These included brief IPIP Interpersonal Circumplex scales (32 items; Markey & Markey, 2009), the 40 strongest correlates of the big two from the IPIP (Saucier, Thalmayer, Payne, et al., 2014), the 6 Big Two scored from the Questionnaire Big Six (QB6; 14 items; Thalmayer & Saucier, 2014), Stability and Plasticity (40 items; DeYoung, 2010), and the SAPA Personality Inventory (SPI; 135 items). The SPI is a hierarchical personality assessment framework, allowing for assessment of Big Five Agreeableness, Extraversion, Conscientiousness, Neuroticism, and Openness based on mean responses of up to 14 items for each trait, plus 27 more narrowly defined traits scored using item-response theory-based parameters for each five-item scale.

Participants were given the option of providing information about 15 behaviors and psychographic characteristics, which are detailed in Table 8 with preregistered hypotheses relating to them. We also queried body mass index, derived from height and weight; discipline of college/university major; occupational field based on the Occupational Information Network taxonomy (U.S. Department of Labor); and occupational prestige, but we did not make a priori hypotheses for these variables.

Analyses

Using mean scores for CCB2I Communion/SSR (23 items) and Agency/Dynamism scales (22 items), correlations with the external personality, behavior, and psychographic variables are reported as point estimates of the Pearson correlations with 95% confidence intervals. Because several of the IPIP-based scales share items, correlations were adjusted to prevent inflated estimates using scoreOverlap in the psych package (Revelle, 2022). Specifically, correlations from overlapping items are replaced by the average correlation for that item. Internal consistency is evaluated using Cronbach's α and general factor saturation with ω_h.

For occupations and majors, response categories used by 100 or more respondents were used. Mean Communion/SSR and Agency/Dynamism scores were calculated for each category, then scores on each trait were ranked across majors or occupations, and a percentile score was computed, indicating the rank position of major/occupation on that personality scale. The percentile scores for each category were then compared between the two scales (Agency/Dynamism percentile score minus Communion/SSR score), to identify majors/occupations most differentially associated with the traits.

Study 4: Results

Psychometric properties of the CCB2I and external scales for the combined samples are reported in Table 9. Internal consistency was high, and factor saturation suggested moderate unidimensionality for both CCB2I scales. The correlation between the CCB2I scales was smaller than in Study 3 ($r = .33$). Reliability values were above .80 for most personality scales, with the exception of QB6 Dynamism and Big Five Openness. It was also low for several four-item interpersonal circumplex (IPC) scales. Associations with the IPC are shown, and geometric coordinates are provided in Figure 2 (following Zimmermann & Wright, 2017).

Zero-order correlations with other variables are displayed in Figure 3, with numeric values provided in the SAPA additional online Tables S26 through S30 accessible at https://osf.io/89x2q/. Partial correlations are in the attached Supplemental Table S7 and

Table 8
SAPA Project Behaviors and Psychographic Characteristics Noting Preregistered Hypotheses

Variable	Communion/SSR	Agency/Dynamism
Age	+	
Gender = female	+	−
Educational attainment level Eight options: "less than 12 years" to "graduate/professional degree"		+
Self-reported health Five options: "poor" to "excellent"	+	
Exercise frequency Six options: "very rarely or never" to "more than 5 times a week"	+	+
Smoking frequency Nine options: "never in my life" to "more than 20 times a day."	−	+
Caffeine consumption Per day, estimated in milligrams	−	
Sleep quality Five options: "excellent" to "poor"	+	
Credit score Five options: "very poor (300 to 579)" to "exceptional (800 to 850)" plus "I don't know."	+	
Volunteering frequency Nine options: "never" to "more than once a week."	+	
Frequency of attending religious services Eight options: "never or rarely" to "every day."	+	
Frequency of Twitter usage Six options: "less than once a month" to "more than a few times a day."		+

Note. + indicates expected positive correlation; − indicates expected negative correlation. SAPA = Synthetic Aperture Personality Assessment; SSR = social self-regulation.

shown visually in Supplemental Figure S2. For two-factor scales (Figure 3a), associations were generally strong and as expected. The other CCB2I scale also tended to be positively correlated (less so for partial correlations), but the associations were stronger for the matching scale. This was also true for Plasticity, although Stability was equally correlated with both CCB2I scales. With the Big Five, correlations were largely as expected from studies of higher order factors, with Communion/SSR moderately positively associated with Agreeableness and Conscientiousness and Agency/Dynamism with Extraversion, Openness, and (negatively) Neuroticism. The SPI-27 scales that associated most strongly with Communion/SSR (in both zero-order and partial correlations) were Compassion, Honesty, and Authoritarianism versus Irritability and Impulsivity. The strongest associations with Agency/Dynamism were for Charisma, Well- Being, Humor, and Sociability versus Anxiety.

The only substantial zero-order associations of the CCB2I scales with sociodemographic and behavioral variables (Figure 3b) were for Agency/Dynamism with health and exercise frequency; these were also positive but low for Communion/SSR. In all other cases correlations were in the expected direction but small in magnitude, with the exception of Twitter usage, which had a small negative correlation with Agency/Dynamism. Agency/Dynamism had a slightly higher correlation with many health variables than did Communion/SSR.

The six majors with the highest Communion/SSR over Agency/Dynamism ratio were (in descending order): Kindergarten/Preschool Education, Community or Social Service, Botany, Graphic Arts, Natural Sciences, and Sociology (see the SAPA additional online Table S33 at https://osf.io/89x2q/). The majors most differentially associated with Agency/Dynamism were Aerospace Engineering, Fashion, Communications and Media Studies, Mathematics Education, Marketing, and Radio/Television/Film Communication (see the SAPA additional online Table S34 at https://osf.io/89x2q/). The occupations most differentially associated with Communion/SSR were Executive Secretary/Administrative Assistant, Registered Nurse, Auditor, Administrative Services Manager, Computer Software Engineer, and Medical Assistant (see the SAPA additional online Table S31 at https://osf.io/89x2q/); those most differentially associated with Agency/Dynamism were Marketing Manager, Human Resources Manager, Lawyer, Community/Social Service Specialist, Manager, and Food Server (see the SAPA additional online Table S32 at https://osf.io/89x2q/).

Study 4: Discussion

Psychometric properties for the CCB2I were good in the SAPA samples. Associations with other big two, Big Five and SPI-27 scales help clarify CCB2I scale coverage. CCB2I scales function most similarly to the other two-factor scales that are based on cross-cultural evidence (IPIP, QB6), but less so with the higher order factors of the Big Five, where Plasticity (Extraversion and Openness) indeed associated more highly with Agency/Dynamism, but Stability (Agreeableness and Conscientiousness) was equally correlated with both CCB2I scales. Associations with the Big Five clarify the misfit: low Neuroticism is included on Stability, but in our results, the association was stronger for Agency/Dynamism. This replicates our results in Study 3, where CCB2I Agency/Dynamism, with its component of positive mood, better predicted lower internalizing tendencies, and Communion/SSR, with its component of even temper, better predicted lower externalizing.

The fit with IPC scales suggests a prototypical Communion/SSR profile (high R^2), with a high degree of variability (amplitude) that is reduced when using partial correlations, modest average correlations across scales (elevation), and projection into the warm/warm submissive part of the interpersonal circle (displacement).

Table 9
Psychometric Properties of the 45-Item CCB2I and Other Scales in the SAPA Data

Scale	Items	*M*	*SD*	α	ω_h	*r*
CCB2I						.33
Communion/SSR	23	4.56	0.92	.87	.63	
Agency/Dynamism	22	3.67	1.08	.91	.57	
IPIP Big Two						.23
SSR	20	4.26	1.13	.86	.72	
Dynamism	20	3.41	1.25	.89	.60	
QB6 Big Two						.21
SSR	7	3.83	1.29	.80	.55	
Dynamism	7	3.93	1.18	.67	.74	
Higher Order Big Five						.31
Stability	24	3.34	1.18	.88	.75	
Plasticity	16	3.95	1.08	.91	.63	
IPIP IPC						
Assured–Dominant (PA)	4	3.27	1.52	.89	.69	
Arrogant–Calculating (BC)	4	3.32	1.45	.89	.57	
Cold-Hearted (DE)	4	3.43	1.39	.71	.64	
Aloof–Introverted (FG)	4	3.96	1.53	.65	.37	
Unassured–Submissive (HI)	4	4.07	1.39	.39	.70	
Unassuming–Ingenuous (JK)	4	4.41	1.20	.74	.35	
Warm–Agreeable (LM)	4	4.53	1.21	.36	.12	
Gregarious–Extraverted (NO)	4	3.12	1.53	.33	.59	
Big Five						
Extraversion	14	3.17	1.11	.86	.63	
Openness	14	4.57	0.82	.77	.70	
Conscientiousness	14	3.95	0.97	.82	.53	
Agreeableness	14	4.23	0.86	.80	.54	
Neuroticism	14	4.28	1.04	.86	.69	
SPI-27	135					
See Supplemental Material	5 each					

Note. *r* = intercorrelation between scales for two-factor models; CCB2I = Cross-Cultural Big Two Inventory; SSR = Social Self-Regulation; QB6 = Questionnaire Big Six; IPIP = International Personality Item Pool; IPC = Interpersonal Circumplex Scales; SPI = SAPA Personality Inventory; SAPA = Synthetic Aperture Personality Assessment.

For Agency/Dynamism, the profile became more prototypical using partial correlations with projection into the warm/warm dominant part of the IPC. These results suggest that the CCB2I scales are rotated somewhat to the right and thus more like Extraversion and Agreeableness compared to the assumptions of the Agency and Communion scales in an IPC framework (e.g., Pincus et al., 2020; Zimmermann & Wright, 2017), at least relative to this IPC measure.

The specificity of the SPI-27 scales confirms the CCB2I content clusters defined in Table 1: Communion/SSR included morality, warmth, respect, industriousness, and even temper, which match the correlations with SPI Honesty, Compassion, Trust, Authoritarianism, Industry, Order, and Well-Being and (negatively) Impulsivity and Irritability. Agency/Dynamism included components of competence, courage, sociability, surgency, and positive mood, which map onto correlations with SPI Creativity, Humor, Industry, Intellect, Adaptability, Sociability, Charisma, Attention Seeking, Emotional Expressiveness, Sensation Seeking, Well-Being, and (negatively) Anxiety.

Correlations with external measures were generally in the directions expected based on prior research, but smaller. In the cases of self-reported health and religious engagement, Study 3 results were replicated in magnitude, direction, and differential association between CCB2I scales. Again here, though with results collapsed across countries, we did not replicate Gebauer et al. (2013) in finding Agency/Dynamism to predict lower religious engagement.

A limitation to these international results is the Western-derived measures and constructs used in the study, for example, the career choices that are not as available and relevant in all countries sampled, and activities like volunteering have very different meaning and prevalence in different societies. While global in composition, this study sampled online participants and, thus, relatively globalized individuals. This means that the study has similar strengths (large samples and reasonably high data quality) and limitations (realistically, results can be generalized only to portions of the societies from which the samples are drawn) as other large international studies used in psychology research. Such studies can generate hypotheses and answer specific questions, but future studies should use better contextualized methods and measure to obtain representative samples in the countries of interest and to assess for the outcomes most meaningful in those societies.

General Discussion

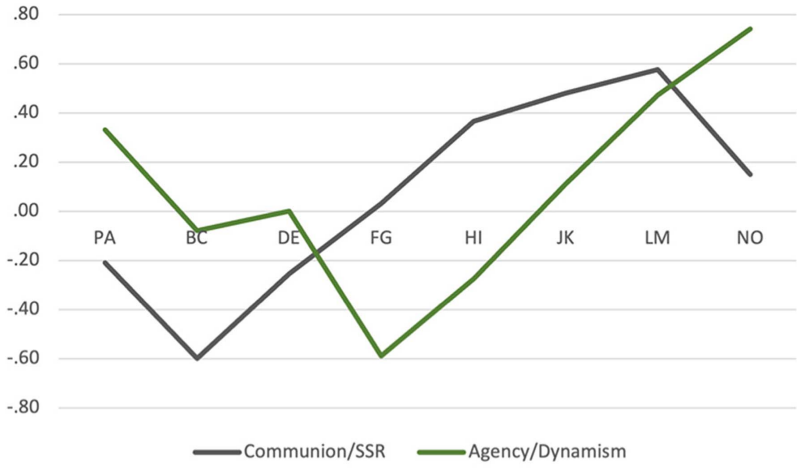
Together, these studies go beyond what has previously been attempted in cross-cultural personality research, using convergent evidence from globally diverse contexts to shape an empirically based, culturally decentered theoretical and measurement model. Commonalities among the content in 11 globally diverse lexical studies led to the identification of 55 trait concepts, which were taken as a foundation for a big two model that is not derivative of prior Big Five work or other theories (Study 1). Items were assessed for cross-cultural utility using data sources from English-speaking contexts around the globe (Study 2). The most reliable and consistent 45 items were then assessed as a big two measurement model and for external validity in community data from three African contexts (Study 3) and in large samples from 13 English-speaking contexts, where they were additionally compared to other personality models to assess convergence/divergence. What these studies teach us about the cross-cultural core of the big two in terms of theory and measurement, and what questions remain, are discussed in turn.

The CCB2 Theoretical Model

Across cultural contexts, there is both great diversity and great regularity in the most salient concepts used to differentiate between people: those encoded through long usage into a single word (Dixon, 1982; Saucier, Thalmayer, & Bel-Bahar, 2014). For example, some languages include many terms to convey fine-grained distinctions within the realm of Extraversion (e.g., L. R. Goldberg, 1990), while others lack such vocabulary but instead allow for precise distinctions around tendencies for hard work versus laziness (Thalmayer et al., 2020) or religious inclination (Thalmayer, Job, et al., 2021). This diversity in denotation gives local flavor to remarkable consistency in connotation, at least at the level of one or two factors. When a single factor is extracted in lexical studies of personality descriptors, this first basic dimension on which people describe each other is easy to recognize as moral evaluation (Saucier, Thalmayer, Payne, et al., 2014). A pole from virtue to vice includes a rich diversity of local terms, with those indicating socially valued, ethically and morally relevant traits tending to have the highest magnitude loadings (Thalmayer, Job, et al., 2021).

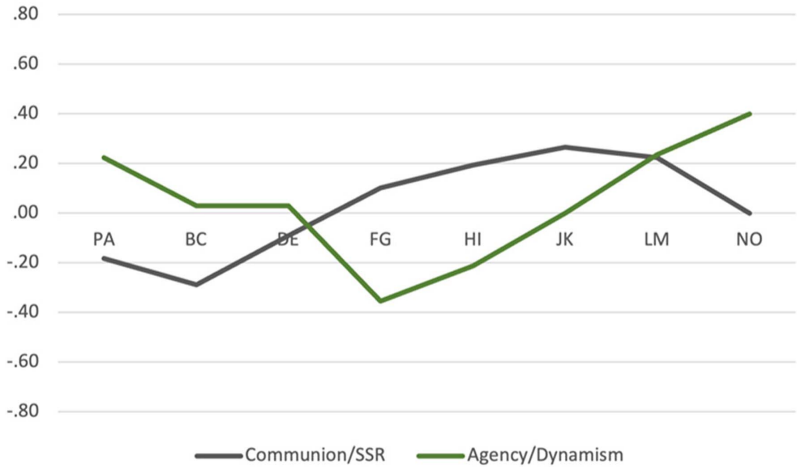
Figure 2

The CCB2I Scales in the Interpersonal Circumplex



(a) Zero-order correlations

Parameter	C/SSR	A/D
Elevation	.07	.16
Amplitude	.53	.44
Displacement	323.29	29.13
R^2	.96	.70



(b) Partial correlations

Parameter	C/SSR	A/D
Elevation	.03	.04
Amplitude	.26	.30
Displacement	307.20	54.14
R^2	.97	.89

Note. Part (a) uses regular correlations for the CCB2I scales, while (b) uses partial correlations. CCB2I = Cross-Cultural Big Two Inventory; Interpersonal Circumplex coordinates: PA = Assured–Dominant; BC = Arrogant–Calculating; DE = Cold-Hearted; FG = Aloof–Introverted; HI = Unassured–Submissive; JK = Unassuming–Ingenuous; LM = Warm–Agreeable; NO = Gregarious–Extraverted; C/SSR = Communion/Social Self-Regulation; A/D = Agency/Dynamism. See the online article for the color version of this figure.

Extracting two factors narrows the first-factor content, honing it to internalization versus rejection of social and cultural codes. The common denominator content of the first factor of two across 11 highly diverse lexical studies, that is morality, warmth, respect, industriousness, and even temper, all share the connotation of moral evaluation or Social Self-Regulation. The second highly robust factor includes competence, bravery, sociability, surgency, and positive mood, all of which share a connotation of Dynamism, the relative proportion of approach versus avoidance tendencies or social effects from interesting to boring.

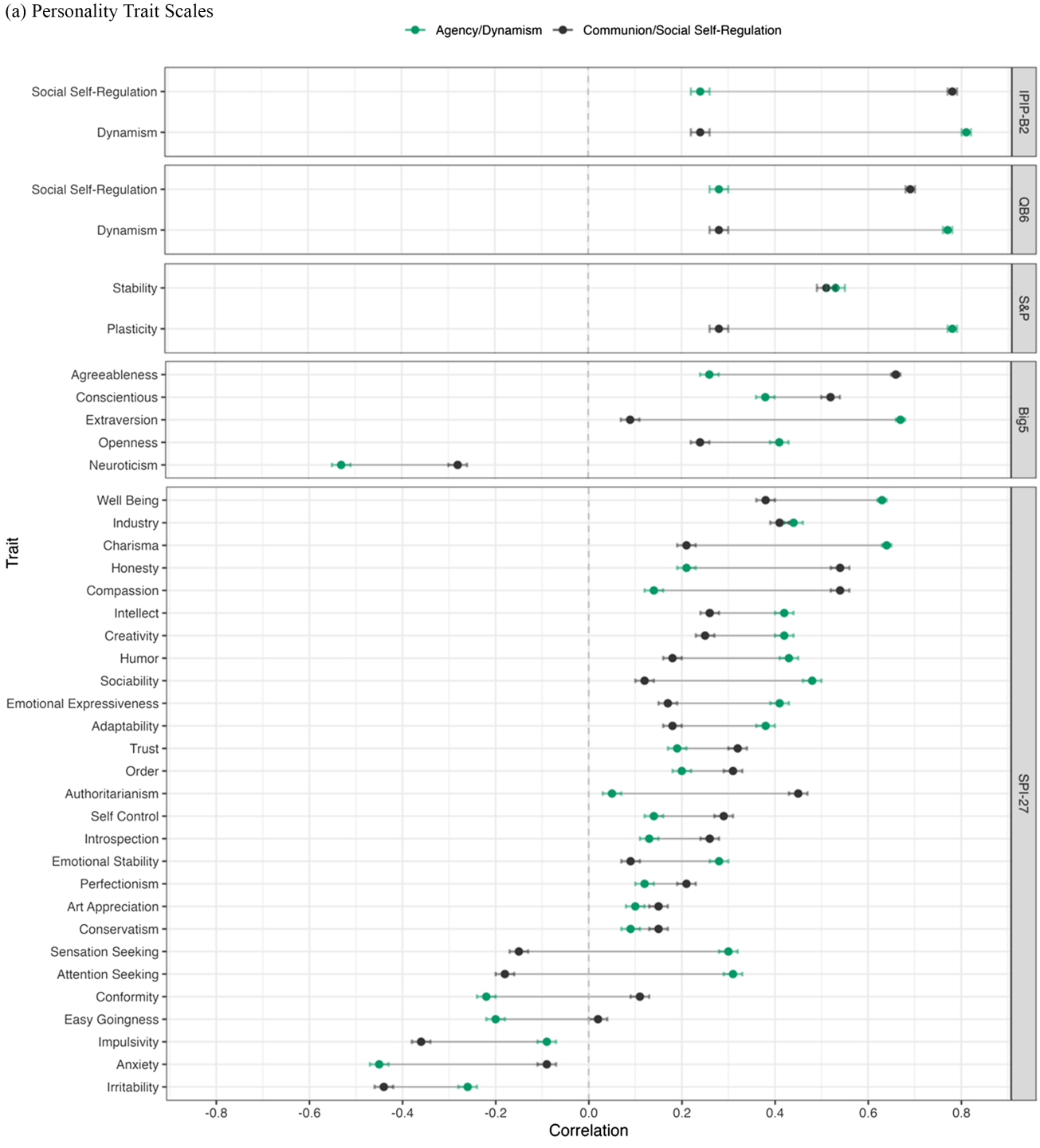
This empirically derived CCB2 captures two driving factors in human variation: a normative code oriented toward coordination with others and a strategic code oriented toward achievement (Bailey, 2018). The model aligns with Bakan's (1966) Communion and Agency as fundamental factors in human society, R. J. Hogan and Robert's (2000) socio-analytic theory, Adler's (1938) social interest and superiority striving, Wiggins's (1979) nurturance and dominance, R. Hogan's (1982) getting along and getting ahead, and McAdams's (1988) power and intimacy motives. It also aligns with two-factor models of stereotypes, person perception and judgment, and impression formation (Wojciszke, 2005), cultural values (Markus & Kitayama, 1991), higher order factors of the Big Five (DeYoung, 2006; Digman, 1997; Jang et al., 2006), and to a degree with the overarching spectra structure of psychopathology (Conway et al., 2019; Kotov et al., 2017). It is also compatible with contemporary integrative interpersonal theory (Hopwood et al., 2021), which seeks to integrate normative personality traits with psychopathology and which similarly specifies that it is the social–adaptive context more than biology that creates these consistent lines of variation. The CCB2 model and inventory thus offer the possibility of integrating a large Western knowledge base into a model built from global evidence.

In Study 4, we observed convergence for the empirically derived CCB2I with three other big two inventories and SPI-27 scales that match CCB2I content clusters. Studies 3 and 4 also helped us to

Figure 3

Correlations Between the CCB2I and External Scales and Variables

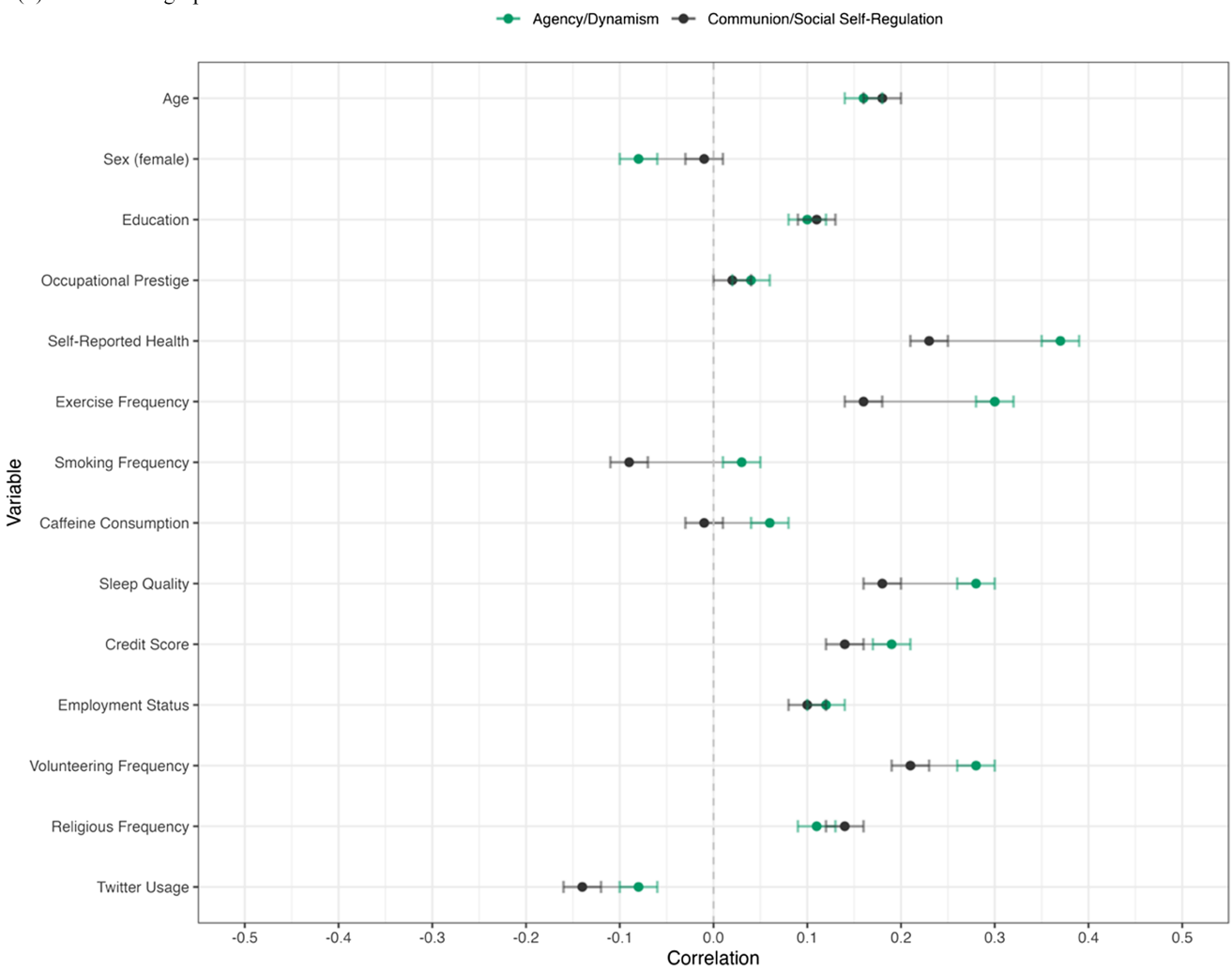
(a) Personality Trait Scales



(*figure continues*)

Figure 3 (*continued*)

(b) Socio-Demographic Variables



Note. CCB2I = Cross-Cultural Big Two Inventory; IPIP-B2 = International Personality Item Pool–Big Two; QB6 = Questionnaire Big Six; S&P = Stability and Plasticity; SPI = SAPA Personality Inventory. See the online article for the color version of this figure.

consider ways the CCB2 diverges from related models. An important distinction has to do with the content typically associated with Big Five Neuroticism. The CCB2I model divides this content into positive mood (Agency/Dynamism) and even temper (Communion/SSR), leading to the distinction observed between lower internalizing and lower externalizing tendencies, respectively (Study 3). In Study 4, Big Five Neuroticism was more associated with low Agency/Dynamism than with low Communion/SSR, which is logical considering that its content is mostly related to internalizing tendencies. This is true of many Big Five inventories, though the NEO inventory platform includes Impulsiveness and Angry Hostility on this factor alongside Anxiety, Depression, Self-Consciousness, and Vulnerability (McCrae & Costa, 1997), which means that distinction between internalizing and externalizing tendencies are possible only at the facet and not at the factor level. This would lead to a lack of differentiation in associations with the CCB2I scales, as seen in Study 4 with Stability and Plasticity (DeYoung, 2006). The Stability scale is predominantly made up of externalizing-related items, secondarily of internalizing content, leading to the lack of differentiation for CCB2I scales. We argue that the distinction provided by the CCB2 model in the domain of mental health is both a conceptual and a practical advantage.

A potential limitation of the CCB2 theoretical and measurement model is the intercorrelation of the two scales. Both contain socially and culturally valued content at one end and undesirable content at the other, so they have a shared evaluative valence component (also discussed by Abele & Wojciszke, 2014). Despite conceptual orthogonality (e.g., Abele & Wojciszke, 2014), intercorrelation is common among two-factor measures. In the SAPA data (Study 4), the .33 intercorrelation of the CCB2I scales was matched by that of Stability and Plasticity and only somewhat larger than those of the QB6 and IPIP Big Two scales (.21–23). Our observed intercorrelations there and in the ALLS data (.53) are comparable to those

reported in other studies, assembled in Supplemental Table S8. Reported big two intercorrelations range from lows of .10 for Trapnell and Paulhus's (2012) agentic and communal life goals (25 items; 272 American students) to highs of .58 and .61 for 10-item (317 Americans, Buchanan & Bardi, 2015; 83 peer ratings, Wojciszke et al., 2009) and 12-item measures of agentic and communal traits (162 peer ratings; Wojciszke & Abele, 2008).

The high correlation between Agency/Dynamism and Communion/SSR in the ALLS samples (Study 3) may capture an elevated alignment in the social desirability of these traits, given the pairing of agentic and communal values in African contexts (e.g., Adjei, 2019). Recent work in ALLS samples including qualitative interviews on ubuntu values and emerging adulthood, and quantitative results on emotional expression all point to the interwoven nature of agentic striving for success and leadership with explicit communal motivations to benefit one's family and community (e.g., Osei-Tutu et al., 2024; Rotzinger et al., 2024). Luckily, the high intercorrelation was successfully addressed using an acquiescence-adjusted version of the scale, which reduced the value to .16, at the low end of all published values.

The CCB2 Measurement Model

In addition to clarifying the theoretical model for the Cross-Cultural Big Two, we sought to establish a culturally decentered measurement model, one drawing on evidence from, in order to work consistently across, global English-speaking contexts. This provides a strong basis for future translations, with further refinement of the inventory to be based on cross-cultural results. The 45-item CCB2I had acceptable psychometric properties and structural validity in our community-based samples of young adults in three African countries and in large online samples from 13 English-speaking countries around the globe.

Measurement invariance was tested in Study 3 and established at the metric level, which allows for comparison of associations across contexts but not mean scores. Tests of measurement invariance were determined to be out of scope for Study 4, due to the complexity of testing it across 13 samples with the SAPA Project's planned-missingness characteristics and to our unwillingness to overfit the scale at this stage of scale development, before translation efforts have had the opportunity to draw from a larger pool of terms. In cross-cultural assessment, measurement invariance is considered a gold standard (e.g., Byrne & Van De Vijver, 2010; Fischer & Karl, 2019). Others have questioned its practicality and utility, pointing out ways that arbitrary thresholds and unmeetable standards inhibit cross-cultural research (Funder & Gardiner, 2024) and that the tests require highly specific configurations of between-group differences in means, limited to the middle ranges of scales (Welzel et al., 2023). We view the testing of measurement invariance to be an important tool in cross-cultural psychology; without it, comparisons across contexts are unreliable and can lead to unsubstantiated conclusions. However, we fully agree that blind adherence to benchmarks from other domains (as argued by Hopwood & Donnellan, 2010), and the practice of treating measurement invariance as a simple yes/no question inhibit cross-cultural psychology. Instead, such tests can best be used in a nuanced and exploratory way, to identify items that work more or less well across contexts, as a part of the process of developing cross-culturally applicable scales (e.g., Thalmayer & Saucier, 2014).

In the current project, the steps that could be taken to improve measurement invariance metrics would mean reducing items, leading to very narrow scales, which would be contrary to our goals at this stage of model development. We prefer to leave the item pool broad in service to future translators or those working in diverse cultural contexts where some aspects may be more salient and useful than others. Some fit indices (e.g., SRMR) reward saturation of content, which we find unsuitable to emphasize, as the big two are by definition very broad traits. Future projects, however, could and should find ways to address this limitation, to be able to compare scores across samples. One idea for establishing measurement invariance while retaining content breadth could be to use large language models with the 45 items as seed terms, to identify related terms to add to the scale, creating a set of items with balanced keying and symmetry in terms of low but consistent intercorrelations (as in Clark & Watson, 1995). Ideally, this could be pursued in multiple languages simultaneously. A large set of diverse items should predict outcomes most efficiently if they have low correlations, with each item only having the latent factor in common. Alternatively, the lumpiness that leads to high SRMR values could be managed by parceling items into clusters.

Other Limitations and Future Directions

Personality psychologists have recently pointed out the limitations of Big Five traits in prediction: Narrower indicators are more precise (e.g., Mõttus, 2016). A two-factor model unfortunately has even less precision than the Big Five. Thus, it is important to emphasize that the CCB2I is not intended as an end point for personality psychology but, instead, as a cross-cultural starting point. This is a framework that provides cross-cultural comparability and that can organize some culture-specific higher dimensionality, for example, the Big Five traits and their facets. Other traits, however, will have little correlation with either of the big two (e.g., SPI Art Appreciation and Conservatism, Figure 3a). An important future direction is to explore and test the extent to which more elaborated local models can be organized as facets under this CCB2I and which are separate from it and to assess their incremental predictive validity. A related idea would be to go beyond these two most replicable factors, taking a broader and less prestructured set of descriptors from as many languages as possible and translating them into a single item pool, for a maximally complete emic–etic exploration.

Another limitation of the present study was the practical need to limit the project to English. Given the work of the authorship team in highly diverse English-speaking contexts, we are acutely aware of the ways English language differs around the globe. We found it a necessary first step for a cross-cultural inventory to take World Englishes and the realities of multilingualism and second-language speakers into consideration. This made translation into other languages out of scope for a single article. However, it is a keen interest of the team, and we are pleased that a translation into an indigenous Latin American language (Tsimane) has already been prepared. Translation of the scales into other languages is a priority. Future work should also assess the functionality of the inventory in administration by oral interview (as in Tsimane), which can allow people in many contexts to respond in their most familiar spoken language rather than the language of schooling.

We would also like to see qualitative approaches used to assess for similarities and differences in the familiarity and prototypical examples given for the terms across contexts, to insure comparability in meaning and valuation (Uugwanga et al., 2024). Relatedly, assessment of the items' relative social desirability across contexts could index potentially varying degrees of valuation of the two domains, their subcomponents, or specific terms. While there is clear evidence for the cross-cultural relevance and similar valence of these traits, it is important to take into consideration any nuances in their relative degree of valuation across contexts. For example, high Agency might be especially valued in competitive contexts, and low Communion/SSR might be especially problematic in tighter cultures and more tolerated in looser ones (e.g., Gelfand et al., 2011).

Conclusions

The four studies of this project enact a creative solution to addressing colonialism in scientific research generally and in personality psychology specifically. We used convergent evidence from around the world to create a culturally decentered model for the study and assessment of personality traits, meaning that the basis of the theoretical and measurement model is international from the outset and does not prioritize a single context. Commonalities in the content of 11 globally diverse lexical studies led to identification of 55 trait concepts (Study 1), which were honed using several data sources to the 45 most consistently used across global English-speaking contexts (Study 2). In the CCB2, Communion/Social Self-Regulation captures the internalization of versus resistance to the normative codes of one's society, and Agency/Dynamism captures approach versus avoidance tendencies. The refined 45-item CCB2I was administered to community samples of 18-year-olds in three African countries (Study 3) and large online samples of adults in 13 English-speaking countries (Study 4) to assess measurement properties, external validity, and convergence with Western-derived two-factor personality models. A difference from some is that content related to psychological disorder tendencies is divided, with Communion/Social Self-Regulation indexing even temper and low externalizing tendencies and Agency/Dynamism indexing positive mood and low internalizing tendencies. The CCB2I model is at a broad level and thus less predictive than narrower traits, and it should be further refined when a larger evidence base is available, to achieve optimal measurement characteristics. It does, however, already offer better suitability for cross-cultural research than other currently available measures. The CCB2I is a culturally decentered starting point from which to build and within which to organize other findings in the study of personality psychology that takes human diversity more fully into account.

References

Abele, A. E., & Hauke, N. (2020). Comparing the facets of the big two in global evaluation of self versus other people. *European Journal of Social Psychology*, *50*(5), 969–982. https://doi.org/10.1002/ejsp.2639

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., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, *38*(7), 1202–1217. https://doi.org/10.1002/ejsp.575

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in experimental social psychology* (Vol. 50, pp. 195–255). Elsevier. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Adams, G., Dobles, I., Gómez, L. H., Kurtiş, T., & Molina, L. E. (2015). Decolonizing psychological science: Introduction to the special thematic section. *Journal of Social and Political Psychology*, *3*(1), 213–238. https://doi.org/10.5964/jspp.v3i1.564

Adjei, S. B. (2019). Conceptualising personhood, agency, and morality for African psychology. *Theory & Psychology*, *29*(4), 484–505. https://doi.org/10.1177/0959354319857473

Adler, A. (1938). *Social interest: A challenge to mankind*. Faber & Faber.

Allik, J., Church, A. T., Ortiz, F. A., Rossier, J., Hřebíčková, M., De Fruyt, F., Realo, A., & McCrae, R. R. (2017). Mean profiles of the NEO personality inventory. *Journal of Cross-Cultural Psychology*, *48*(3), 402–420. https://doi.org/10.1177/0022022117692100

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

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. https://doi.org/10.1037/0022-3514.86.2.356

Bailey, F. G. (2018). *Stratagems and spoils: A social anthropology of politics* (1st ed.). Routledge. https://doi.org/10.4324/9780429497094

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. RandMcNally.

Bleidorn, W., Klimstra, T. A., Denissen, J. J. A., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2013). Personality maturation around the world: A cross-cultural examination of social-investment theory. *Psychological Science*, *24*(12), 2530–2540. https://doi.org/10.1177/0956797613498396

Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The Family Affluence Scale as a measure of national wealth: Validation of an adolescent self-report measure. *Social Indicators Research*, *78*(3), 473–487. https://doi.org/10.1007/s11205-005-1607-6

Brewer, M. B., & Chen, Y.-R. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological Review*, *114*(1), 133–151. https://doi.org/10.1037/0033-295X.114.1.133

Buchanan, K., & Bardi, A. (2015). The roles of values, behavior, and value-behavior fit in the relation of agency and communion to well-being. *Journal of Personality*, *83*(3), 320–333. https://doi.org/10.1111/jopy.12106

Byrne, B. M., & Van De Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132. https://doi.org/10.1080/15305051003637306

Central Intelligence Agency. (2021). *The world factbook 2021*.

Cheung, F. M., Cheung, S. F., & Fan, W. (2013). From Chinese to cross-cultural personality inventory: A combined emic–etic approach to the study of personality in culture. In M. J. Gelfand, C. Chiu, & Y. Hong (Eds.), *Advances in culture and psychology: Volume 3* (pp. 117–179). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199930449.001.0001

Church, A. T., Alvarez, J. M., Mai, N. T. Q., French, B. F., Katigbak, M. S., & Ortiz, F. A. (2011). Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet functioning in the revised NEO Personality Inventory. *Journal of Personality and Social Psychology*, *101*(5), 1068–1089. https://doi.org/10.1037/a0025290

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Condon, D. M., Roney, E., & Revelle, W. (2017). A SAPA project update: On the structure of phrased self-report personality items. *Journal of Open Psychology Data*, *5*, Article 3. https://doi.org/10.5334/jopd.32

Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., Mullins-Sweatt, S. N., Shackman, A. J., Skodol, A. E., South, S. C., Sunderland, M., Waszczuk, M. A., Zald, D. H., Afzali, M. H., Bornovalova, M. A., Carragher, N., Docherty, A. R., Jonas, K. G., Krueger, R. F., … Eaton, N. R. (2019). A hierarchical taxonomy of psychopathology can transform mental health research. *Perspectives on Psychological Science*, *14*(3), 419–436. https://doi.org/10.1177/1745691618810696

Costa, P. T., & McCrae, R. R. (2014). The NEO Inventories1. In R. P. Archer & S. R. Smith (Eds.), *Personality assessment* (pp. 229–260). Routledge.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(1), Article 7. https://doi.org/10.7275/JYJ1-4868

Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the health behaviour in school-aged children (HBSC) family affluence scale. *Social Science & Medicine*, *66*(6), 1429–1436. https://doi.org/10.1016/j.socscimed.2007.11.024

Davies, M. (2008). *The corpus of contemporary American English (COCA): 560 million words, 1990–present*. https://www.english-corpora.org/coca/

De Raad, B., Barelds, D. P. H., Timmerman, M. E., De Roover, K., Mlačić, B., & Church, A. T. (2014). Towards a pan–cultural personality structure: Input from 11 psycholexical studies. *European Journal of Personality*, *28*(5), 497–510. https://doi.org/10.1002/per.1953

DeSalvo, K. B., Fan, V. S., McDonell, M. B., & Fihn, S. D. (2005). Predicting mortality and healthcare utilization with a single question. *Health Services Research*, *40*(4), 1234–1246. https://doi.org/10.1111/j.1475-6773.2005.00404.x

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. (2010). Toward a theory of the big five. *Psychological Inquiry*, *21*(1), 26–33. https://doi.org/10.1080/10478401003648674

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the big five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, *33*(4), 533–552. https://doi.org/10.1016/S0191-8869(01)00171-4

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

Dixon, R. M. (1982). *Where have all the adjectives gone?: And other essays in semantics and syntax*. De Gruyter Mouton. https://doi.org/10.1515/9783110822939

Dowd, C. (2020). *Two samples: Fast permutation based two sample tests*. R package Version 1.0. 0. [Computer software].

Dowd, C. (2022). *twosamples: Fast permutation based two sample tests* [Computer software]. https://CRAN.R-project.org/package=twosamples

Fetvadjiev, V. H., Meiring, D., van de Vijver, F. J. R., Nel, J. A., & Hill, C. (2015). The South African Personality Inventory (SAPI): A culture-informed instrument for the country's main ethnocultural groups. *Psychological Assessment*, *27*(3), 827–837. https://doi.org/10.1037/pas0000078

Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, *10*, Article 1507. https://doi.org/10.3389/fpsyg.2019.01507

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

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

Frimer, J. A., Walker, L. J., Dunlop, W. L., Lee, B. H., & Riches, A. (2011). The integration of agency and communion in moral personality: Evidence of enlightened self-interest. *Journal of Personality and Social Psychology*, *101*(1), 149–163. https://doi.org/10.1037/a0023780

Funder, D. C., & Gardiner, G. (2024). MIsgivings about measurement invariance. *European Journal of Personality*. Advance online publication. https://doi.org/10.1177/08902070241228338

Gebauer, J. E., Paulhus, D. L., & Neberich, W. (2013). Big two personality and religiosity across cultures: Communals as religious conformists and agentics as religious contrarians. *Social Psychological & Personality Science*, *4*(1), 21–30. https://doi.org/10.1177/1948550612442553

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Subirats Ferrer, M., Fischlmayr, I. C., … Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100–1104. https://doi.org/10.1126/science.1197754

Goldberg, L. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (pp. 141–165). SAGE Publication.

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

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11), S78–S94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

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

Henrich, J. (2020). *The WEIRDest people in the world: How the west became psychologically peculiar and particularly prosperous*. Penguin.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hogan, R. (1982). A socioanalytic theory of personality. In E. Page (Ed.), *Nebraska symposium on motivation* (Vol. 30, pp. 55–89). Universtiy of Nebraska Press.

Hogan, R., & Cheek, J. M. (1983). Identity, authenticity, and maturity. In T. R. Sarbin & E. Scheibe (Eds.), *Studies in social identity* (pp. 339–357). Praeger.

Hogan, R. J., & Roberts, B. W. (2000). A socioanalytic perspective on person–environment. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person–environment psychology: New directions and perspectives* (2nd ed., pp. 1–24). Lawrence Erlbaum.

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social*

Psychology Review, *14*(3), 332–346. https://doi.org/10.1177/1088868310361240

Hopwood, C. J., Pincus, A. L., & Wright, A. G. C. (2021). Six assumptions of contemporary integrative interpersonal theory of personality and psychopathology. *Current Opinion in Psychology*, *41*, 65–70. https://doi.org/10.1016/j.copsyc.2021.03.007

Irwing, P., Hughes, D. J., Tokarev, A., & Booth, T. (2024). Towards a taxonomy of personality facets. *European Journal of Personality*, *38*(3), 494–515. https://doi.org/10.1177/08902070231200919

Jang, K. L., Livesley, W. J., Ando, J., Yamagata, S., Suzuki, A., Angleitner, A., Ostendorf, F., Riemann, R., & Spinath, F. (2006). Behavioral genetics of the higher-order factors of the Big Five. *Personality and Individual Differences*, *41*(2), 261–272. https://doi.org/10.1016/j.paid.2005.11.033

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*(6), 899–913. https://doi.org/10.1037/0022-3514.89.6.899

Kendler, K. S., Aggen, S. H., Knudsen, G. P., Røysamb, E., Neale, M. C., & Reichborn-Kjennerud, T. (2011). The structure of genetic and environmental risk factors for syndromal and subsyndromal common DSM-IV axis I and all axis II disorders. *The American Journal of Psychiatry*, *168*(1), 29–39. https://doi.org/10.1176/appi.ajp.2010.10030340

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry*, *60*(9), 929–937. https://doi.org/10.1001/archpsyc.60.9.929

Kirkpatrick, A. (2012). English as an international language in Asia: Implications for language education. In A. Kirkpatrick & R. Sussex (Eds.), *English as an international language in Asia: Implications for language education* (Vol. 1, pp. 29–44). Springer. https://doi.org/10.1007/978-94-007-4578-0_3

Koenig, H., Parkerson, G. R., Jr., & Meador, K. G. (1997). Religion index for psychiatric research. *The American Journal of Psychiatry*, *154*(6), 885–886. https://doi.org/10.1176/ajp.154.6.885b

Kortmann, B., & Schneider, E. W. (Eds.). (2004a). *A handbook of varieties of English: A multimedia reference tool*. Mouton de Gruyter. https://doi.org/10.1515/9783110197181

Kortmann, B., & Schneider, E. W. (Eds.). (2004b). *A handbook of varieties of English: A multimedia reference tool. Two volumes plus CD-ROM*. De Gruyter. https://doi.org/10.1515/9783110197181

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kretzschmar, W. A. (2010). The development of standard American English. In A. Kirkpatrick (Ed.), *The Routledge handbook of world Englishes* (pp. 96–111). Routledge.

Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The structure and stability of common mental disorders (DSM-III-R): A longitudinal–epidemiological study. *Journal of Abnormal Psychology*, *107*(2), 216–227. https://doi.org/10.1037/0021-843X.107.2.216

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior, and personality: Modeling the externalizing spectrum. In G. A. Marlatt & K. Witkiewitz (Eds.), *Addictive behaviors: New readings on etiology, prevention, and treatment* (pp. 59–88). American Psychological Association. https://doi.org/10.1037/0021-843X.111.3.411

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., Rubio-Codina, M., & Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. *Science Advances*, *5*(7), Article eaaw5226. https://doi.org/10.1126/sciadv.aaw5226

Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). A hierarchical causal taxonomy of psychopathology across the life span. *Psychological Bulletin*, *143*(2), 142–186. https://doi.org/10.1037/bul0000069

Markey, P. M., & Markey, C. N. (2009). A brief assessment of the interpersonal circumplex: The IPIP-IPC. *Assessment*, *16*(4), 352–361. https://doi.org/10.1177/1073191109340382

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Mastor, K. A., Thalmayer, A. G., Swami, V., & Saucier, G. (n.d.). *'Attracting-Heart' or 'Long-Handed'?: The structure of Malay personality terms* [Unpublished manuscript]. School of Liberal Studies, National University of Malaysia.

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

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

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82. https://doi.org/10.1037/1082-989X.7.1.64

Melchers, G., Shaw, P., & Sundkvist, P. (2019). *World Englishes* (3rd ed.). Routledge. https://doi.org/10.4324/9781351042581

Mõttus, R. (2016). Towards more rigorous personality trait–outcome research. *European Journal of Personality*, *30*(4), 292–303. https://doi.org/10.1002/per.2041

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big Few traits. *European Journal of Personality*, *34*(6), 1175–1201. https://doi.org/10.1002/per.2311

Osei-Tutu, A., Salvador, C. E., Thalmayer, A. G., Ishii, K., Wang, E., Naudé, L., Suo, T., Shino, E. N., Rossmaier, A., Asatsa, S., Thomas, C. & Kitayama, S. (2024). *Understanding self-promotive interdependence: An exploration in Sub-Saharan Africa*. PsyArXiv. https://doi.org/10.31234/osf.io/egn8h

Pincus, A. L., Hopwood, C. J., & Wright, A. G. C. (2020). The interpersonal situation: An integrative framework for the study of personality, psychopathology, and psychotherapy. In J. F. Rauthmann, R. A. Sherman, & D. C. Funder (Eds.), *The Oxford handbook of psychological situations* (pp. 123–142). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.4

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research* (version 2.2.9) [Computer software].

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment*, *31*(12), 1395–1411. https://doi.org/10.1037/pas0000754

Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2017). Web and phone based data collection using planned missing designs. In N. G. Fielding, G. Blank, & R. M. Lee (Eds.), *The SAGE handbook of online research methods* (2nd ed., pp. 578–595). SAGE Publications.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rossier, J., Ouedraogo, A., & Dahourou, D. (2017). Personality structure and assessment in French-speaking African cultures. In A. T. Church (Ed.), *The Praeger handbook of personality across cultures: Trait psychology across cultures* (pp. 73–103). Praeger/ABC-CLIO.

Rotzinger, J. S., Jensen, L. A., & Thalmayer, A. G. (2024). *Ubuntu/harambee among young adults in Namibia and Kenya*. https://osf.io/preprints/psyarxiv/cye2s

Saucier, G. (2009). Recurrent personality dimensions in inclusive lexical studies: Indications for a big six structure. *Journal of Personality*, *77*(5), 1577–1614. https://doi.org/10.1111/j.1467-6494.2009.00593.x

Saucier, G. (2010). The structure of social effects: Personality as impact on others. *European Journal of Personality*, *24*(3), 222–240. https://doi.org/10.1002/per.761

Saucier, G., & Iurino, K. (2020). High-dimensionality personality structure in the natural language: Further analyses of classic sets of English-language trait-adjectives. *Journal of Personality and Social Psychology*, *119*(5), 1188–1219. https://doi.org/10.1037/pspp0000273

Saucier, G., Thalmayer, A. G., & Bel-Bahar, T. S. (2014). Human attribute concepts: Relative ubiquity across twelve mutually isolated languages. *Journal of Personality and Social Psychology*, *107*(1), 199–216. https://doi.org/10.1037/a0036492

Saucier, G., Thalmayer, A. G., Payne, D. L., Carlson, R., Sanogo, L., Ole-Kotikash, L., Church, A. T., Katigbak, M. S., Somer, O., Szarota, P., Szirmák, Z., & Zhou, X. (2014). A basic bivariate structure of personality attributes evident across nine languages. *Journal of Personality*, *82*(1), 1–14. https://doi.org/10.1111/jopy.12028

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, *38*(2), 173–212. https://doi.org/10.1177/0022022106297299

Schreier, D., Hundt, M., & Schneider, E. W. (2020). World Englishes: An introduction. In D. Schreier, M. Hundt, & E. W. Schneider (Eds.), *The Cambridge handbook of world Englishes* (1st ed., pp. 1–22). Cambridge University Press. https://doi.org/10.1017/9781108349406.001

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology*, *84*(1), 60–79. https://doi.org/10.1037/0022-3514.84.1.60

Silan, M., Adetula, A., Basnight-Brown, D. M., Forscher, P. S., Dutra, N., & Ijzerman, H. (2021, August 26). Psychological science needs the entire globe, Part 2. Let's talk about the "C" word: Colonialism and the challenges of psychological science in the developing world. *Observer*. https://www.psychologicalscience.org/observer/psychological-science-needs-the-entire-globe-part-2

Tanaka, Y., & Osgood, C. E. (1965). Cross-culture, cross-concept, and cross-subject generality of affective meaning systems. *Journal of Personality and Social Psychology*, *2*(2), 143–153. https://doi.org/10.1037/h0022392

Thalmayer, A. G., Asatsa, S., Shino, E. N., Naudé, L., Laher, S., Hassem, T., Florence, M., Adonis, T.-A., Uugwanga, S. N., Rotzinger, J. S., Hofmann, D., Makunda, J., Botha, C., Murangi, A., & Shirima, C. M. (2024). Lifespan research in Kenya, Namibia, and South Africa: Cohort profile of the Africa long life study. *Personality Science*. Advance online publication. https://doi.org/10.1177/27000710241264492

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., Marshall, J., & Scalise, K. (2023). The International Mental Health Assessment: Validation of an efficient screening inventory. *Collabra. Psychology*, *9*(1), Article 74546. https://doi.org/10.1525/collabra.74546

Thalmayer, A. G., & Saucier, G. (2014). The questionnaire big six in 26 nations: Developing cross–culturally applicable big six, big five and big two inventories. *European Journal of Personality*, *28*(5), 482–496. https://doi.org/10.1002/per.1969

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

Thalmayer, A. G., Saucier, G., & Rotzinger, J. S. (2022). Absolutism, relativism, and universalism in personality traits across cultures: The case of the Big Five. *Journal of Cross-Cultural Psychology*, *53*(7–8), 935–956. https://doi.org/10.1177/00220221221111813

Thalmayer, A. G., Saucier, G., Shino, E. N., & Job, S. (2021). The Khoekhoegowab Personality Inventory: The comparative validity of a locally derived measure of traits. *Frontiers in Psychology*, *12*, Article 694205. https://doi.org/10.3389/fpsyg.2021.694205

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, *76*(1), 116–129. https://doi.org/10.1037/amp0000622

Thielmann, I., Akrami, N., Babarović, T., Belloch, A., Bergh, R., Chirumbolo, A., Čolović, P., de Vries, R. E., Dostál, D., Egorova, M., Gnisci, A., Heydasch, T., Hilbig, B. E., Hsu, K.-Y., Izdebski, P., Leone, L., Marcus, B., Međedović, J., Nagy, J., … Lee, K. (2020). The HEXACO–100 across 16 languages: A large-scale test of measurement invariance. *Journal of Personality Assessment*, *102*(5), 714–726. https://doi.org/10.1080/00223891.2019.1614011

Toscanelli, C., Shino, E., Robinson, S. L., & Thalmayer, A. G. (2022). Religiousness worldwide: Translation of the Duke University Religion Index into 20 languages and validation across 27 nations. *Measurement Instruments for the Social Sciences*, *4*(1), Article 13. https://doi.org/10.1186/s42409-022-00041-2

Trapnell, P. D., & Paulhus, D. L. (2012). Agentic and communal values: Their scope and measurement. *Journal of Personality Assessment*, *94*(1), 39–52. https://doi.org/10.1080/00223891.2011.627968

Uugwanga, S. N., Naudé, L., & Thalmayer, A. G. (2024). *Becoming an Ovambo Adult: Growing into Prosocial Agency in Africa* [Manuscript in preparation]. Department of Psychology, University of Zurich.

Van Aarde, N., Meiring, D., & Wiernik, B. M. (2017). The validity of the Big Five personality traits for job performance: Meta-analyses of South African studies. *International Journal of Selection and Assessment*, *25*(3), 223–239. https://doi.org/10.1111/ijsa.12175

Vollebergh, W. A. M., Iedema, J., Bijl, R. V., de Graaf, R., Smit, F., & Ormel, J. (2001). The structure and stability of common mental disorders: The NEMESIS study. *Archives of General Psychiatry*, *58*(6), 597–603. https://doi.org/10.1001/archpsyc.58.6.597

Welzel, C., Brunkert, L., Kruse, S., & Inglehart, R. F. (2023). Non-invariance? An overstated problem with misconceived causes. *Sociological Methods & Research*, *52*(3), 1368–1400. https://doi.org/10.1177/0049124121995521

Wierzbicka, A. (1996). *Semantics: Primes and universals*. Oxford University Press. https://doi.org/10.1093/oso/9780198700029.001.0001

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*(3), 395–412. https://doi.org/10.1037/0022-3514.37.3.395

Wiggins, J. S., & Broughton, R. (1991). A geometric taxonomy of personality scales. *European Journal of Personality*, *5*(5), 343–365. https://doi.org/10.1002/per.2410050503

Wojciszke, B. (1997). Parallels between competence- versus morality-related traits and individualistic versus collectivistic values. *European*

Journal of Social Psychology, *27*(3), 245–256. https://doi.org/10.1002/(SICI)1099-0992(199705)27:3<245::AID-EJSP819>3.0.CO;2-H

Wojciszke, B. (2005). Morality and competence in person- and self-perception. *European Review of Social Psychology*, *16*(1), 155–188. https://doi.org/10.1080/10463280500229619

Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology*, *38*(7), 1139–1147. https://doi.org/10.1002/ejsp.549

Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. *European Journal of Social Psychology*, *39*(6), 973–990. https://doi.org/10.1002/ejsp.595

Ybarra, O., Chan, E., & Park, D. (2001). Young and old adults' concerns about morality and competence. *Motivation and Emotion*, *25*(2), 85–100. https://doi.org/10.1023/A:1010633908298

Zeldow, P. B., Daugherty, S. R., & McAdams, D. P. (1988). Intimacy, power, and psychological well-being in medical students. *Journal of Nervous and Mental Disease*, *176*(3), 182–187. https://doi.org/10.1097/00005053-198803000-00007

Zimmermann, J., & Wright, A. G. C. (2017). Beyond description in interpersonal construct validation: Methodological advances in the circumplex structural summary approach. *Assessment*, *24*(1), 3–23. https://doi.org/10.1177/1073191115621795

Received March 6, 2024
Revision received July 17, 2024
Accepted August 13, 2024 ■