6

Criteria and Methods for Assessing Cultural Universality of Cognitive Representations Underlying Complex Psychological Constructs

Yiu-Fai Yung, Erica G. Hepper, Tim Wildschut, and Constantine Sedikides

Introduction

Do individuals in different cultures have the same cognitive representation of a given psychological construct? For example, do they agree on the meaning of complex emotions, such as sympathy or nostalgia? This is the daunting universality question posed in cross-cultural research. That complex constructs such as emotions lack explicit formal definitions presents a formidable obstacle to answering this question; as Fehr and Russell (1984) state, "Everyone knows what an emotion is, until asked to give a definition. Then, it seems, no one knows" (p. 464). However, viewing emotions (and other complex constructs) from a prototype perspective suggests possible solutions to this problem (Shaver et al., 1987).

According to prototype theory (Rosch, 1978), knowledge is formed on the basis of repeated experience and becomes organized around a generic representation or prototype of the construct. From this perspective, many cognitive constructs are best conceptualized as fuzzy sets with vague boundaries. Rather than being delimited by necessary and sufficient properties, these fuzzy sets are defined by features that are representative or typical of the construct, with highly representative features occupying a more central place in the prototype. Thus, even if a construct cannot be delineated by sharp boundaries, individuals can report whether a particular feature is relatively central or peripheral to said construct. By harnessing this strategy,

Yiu-Fai Yung, Erica G. Hepper, Tim Wildschut, and Constantine Sedikides, *Criteria and Methods for Assessing Cultural Universality of Cognitive Representations Underlying Complex Psychological Constructs* In: *Handbook of Advances in Culture and Psychology*. Edited by: Michele J. Gelfand, Chi-yue Chiu, and Ying-yi Hong, Oxford University Press.
 DOI: 10.1093/oso/9780197689783.003.0006

prototype methods have shed light on individuals' conceptions of a wide range of emotions, including love, commitment (Fehr, 1988), hate, anger, jealousy (Fitness & Fletcher, 1993), respect (Frei & Shaver, 2002), forgiveness (Kearns & Fincham, 2004), gratitude (Lambert et al., 2009), shame (Hurtado de Mendoza et al., 2010), and vengefulness (Elshout et al., 2017). Beyond the field of emotions, scholars have applied the prototype approach to gain insight into a rich variety of domains, including conceptions of personality types (Cantor & Mischel, 1979; Chaplin et al., 1988), modesty (Gregg et al., 2008; Shi et al., 2020), psychiatric conditions (Horowitz et al., 1982; Westen et al., 2012), social situations (Cantor et al., 1982; Uskul et al., 2014), and social categories (Brewer et al., 1981; Sesko & Biernat, 2010).

In their influential article, Shaver et al. (1987) suggested that the prototype approach could usher in an era of research on cross-cultural similarities and differences in emotion conceptions. They proposed that although individuals from different cultures may have difficulty giving a clear-cut definition of an emotion, they should be able to rate whether a particular feature is relatively representative or unrepresentative of the emotion. These prototypicality ratings could then form the basis of cross-cultural comparisons. Yet, although cross-cultural emotion research has blossomed (for reviews, see Mesquita & Frijda, 1992; Russell, 1991; van Hemert et al., 2007), relatively few scholars have capitalized on the prototype approach in the way Shaver et al. envisaged (cf. Fischer et al., 1999; Hepper et al., 2014; Hurtado de Mendoza et al., 2010). A possible reason for this scarcity is the lack of a theoretical framework to guide the comparison of prototypicality ratings among cultures. For example, confirmatory factor analysis (CFA), which is a popular statistical technique in cross-cultural research, does not lend itself to testing the specific postulates of prototype theory. In addition, CFA procedures that are commonly used to test invariance assumption are restrictive (Funder, 2020; Gardiner et al., 2019) and "can be extremely problematic both statistically and substantively" (Byrne & Van de Vijver, 2010, p. 107). In addition, diverse invariances indices are often applied inconsistently by different researchers and may lack practical significance (Ock et al., 2020). Arguably, this lack of framework has limited cross-cultural research in social and personality psychology because researchers lack the guidance and tools to assess the replicability of cognitive representations across cultures, which in turn perpetuates a reliance on so-called WEIRD samples (i.e., Western, educated, industrialized, rich, and democratic; Henrich et al., 2010).

Our key objective, then, is to propose a framework that consists of practical criteria for assessing cultural universality of prototypes for social psychological constructs. We present a case study of conceptions of "nostalgia" to illustrate how these criteria and the associated methodology can address the cultural universality issues. One important methodological departure from most cross-cultural studies is that we do not emphasize invariance tests via CFA (Byrne & Van de Vijver, 2010; Matsumoto & Van de Vijver, 2011; Millsap, 2011; Van de Vijver & Leung, 1997) in establishing cross-cultural universality. Given that our primary goal is to demonstrate the usefulness of our new methodology (rather than examine critically the CFA approach), we postpone overarching comparisons with the CFA approach to the Discussion section. Another goal is to illustrate the utility of often-ignored exploratory multivariate statistical techniques, such as multidimensional scaling (Kruskal & Wish, 1978) and cluster analysis (Aldenderfer & Blashfield, 1984; Arabie et al., 1987), for studying mean patterns in cultures. We show how these exploratory techniques can help researchers generate insightful hypotheses for further investigations.

We first review the prototype approach to complex constructs, focusing in particular on studies of the nostalgia prototype by Hepper et al. (2012, 2014). These studies set the stage for introducing our four operational criteria for judging cross-cultural universality in multiple populations. We then apply these criteria to the cross-cultural data collected by Hepper et al. (2014) and devise statistical tests for assessing the universality of nostalgia conceptions. In addition, we use multidimensional scaling results to interpret the clustering patterns in Hepper et al. (2014). We conclude by discussing the strengths and limitations of our proposed methodology for studying cross-cultural universality.

Prototype Studies of Nostalgia

To characterize lay conceptualizations of our illustrative case, nostalgia, Hepper et al. (2012) adopted a prototype approach. They proposed that nostalgia is a complex emotion lacking a clear-cut definition and sharp boundaries. For example, defining nostalgia as either positive or negative is simplistic, but positive emotions may be more representative of nostalgic experiences than negative emotions. Moreover, a particular experience does not qualify as either nostalgic or non-nostalgic, but some experiences are more representative of nostalgia than others. Relying on UK and U.S. samples, Hepper et al. found that nostalgia was characterized by 35 features, with some features being more prototypical than others. For example, when asked to rate the relevance of a set of features to the construct “nostalgia” (1 = *not at all related*, 8 = *extremely related*), participants rated features such as “memory/memories,” “feeling/emotion,” and “happiness” much higher than features such as “regret,” “sadness/depressed,” and “lethargy/laziness.” We denote highly prototypical features as central and less prototypical features as peripheral. We provide all 35 features of nostalgia along with descriptive statistics in Table 6.1.

Table 6.1 Means and Standard Deviations of Nostalgia Features in the Normed UK Sample

	N	*N* **Miss**	**Mean**	*SD*
Central 1 (C1)				
Memory/memories	102	0	7.10	1.17
The past	101	1	7.00	1.18
Fond memories	102	0	6.73	1.28
Remembering	101	1	6.63	1.41
Reminiscence	100	2	6.54	1.41
Feeling/emotion	101	1	6.47	1.35
Personal meaning	101	1	6.39	1.68
Longing/yearning	100	2	6.32	1.55
Social relationships	101	1	6.28	1.48
Central 2 (C2)				
Memorabilia/keepsakes	101	1	6.04	1.71
Rose-tinted memory	101	1	6.01	1.62
Happiness	100	2	5.95	1.63
Childhood/youth	101	1	5.88	1.68
Sensory triggers	102	0	5.85	1.61
Thinking	101	1	5.84	1.68
Reliving/dwelling	101	1	5.75	1.82
Missing/loss	101	1	5.70	1.70
Wanting to return to past	102	0	5.68	1.81
Peripheral 1 (P1)				
Comfort/warmth	102	0	5.59	1.65
Wishing/desire	102	0	5.42	1.68
Dreams/daydreaming	102	0	5.33	1.67
Mixed feelings	101	1	5.04	1.94
Change	101	1	4.78	1.80
Calm/relaxed	101	1	4.64	1.66
Regret	102	0	4.33	1.91
Homesickness	101	1	4.06	1.92
Prestige/success	101	1	4.05	1.87
Peripheral 2 (P2)				
Aging/old people	100	2	4.00	2.06
Loneliness	102	0	3.76	1.90
Sadness/depressed	101	1	3.58	1.94
Negative past	102	0	3.33	1.94
Distortion/illusions	102	0	3.30	1.99
Solitude	100	2	3.22	1.64
Pain/anxiety	100	2	3.03	1.84
Lethargy/laziness	102	0	2.46	1.61

To examine whether other cultures have similar conceptions of nostalgia, Hepper et al. (2014) extended their investigation to 18 countries from five continents. After validating the translation of the 35 prototypical features, participants were asked in their own language to rate the relevance of these 35 features to the construct "nostalgia." Hepper et al. concluded that, except for mild departures in some African countries, conceptions of nostalgia are near-universal. We revisit some of the statistical analyses in this work. More important, we use their research as a case study to illustrate our method for assessing cross-cultural universality of complex constructs.

Structural Properties of the Nostalgia Features

According to prototype theory (Rosch, 1978), complex constructs are cognitively represented in terms of features that vary in centrality (vs. peripherality). For example, based on Hepper et al.'s (2012) research, the construct "nostalgia" has 35 defining features (or attributes) that vary in centrality. An implication of prototype theory is that the cross-cultural universality of nostalgia conceptions can be assessed by examining whether these 35 features and their structural properties are preserved in different cultures. In this section, we systematically describe important structural properties of features in prototype theory.

We take the 35 nostalgia features identified by Hepper et al. (2012) as a generic example. Let $A = \{a_1, a_2, \ldots, a_{35}\}$ denote this set of 35 features. Suppose that in a population (i.e., a particular culture or country), all individuals rate these 35 features according to their "relatedness" or "representativeness" (i.e., prototypicality) to the construct "nostalgia" on a rating scale, where larger values represent higher relatedness. These ratings are represented by a set of 35 random variables $x_1, x_2, \ldots, x_{35}$. Let $\mu_1, \mu_2, \ldots, \mu_{35}$ be the population means of the rating and $\sigma_1, \sigma_2, \ldots, \sigma_{35}$ be the standard deviations of the ratings. Without loss of generality, assume that the features are ordered by their prototypicality of nostalgia so that $\mu_1 > \mu_2 > \cdots > \mu_{35}$. Hence, feature a_1 is the most prototypical of nostalgia and a_{35} is the least prototypical of (but still related to) nostalgia. In prototype theory, the more prototypical features, such as a_1, a_2, a_3, are called central features, and the less prototypical features, such as a_{33}, a_{34}, a_{35}, are called peripheral features. Hence, the most pertinent structural property of features is that they are ordered according to their prototypicality. This is stated formally in the following section.

Property 1: Ordering of Features

In the population, the features $A = \{a_1, a_2, a_3, \ldots\}$, of a complex construct are ordered from the most prototypical to the least prototypical according to the average prototypicality rating (i.e., $\mu_1 > \mu_2 > \mu_3 > \cdots$) of the features by all individuals.

Although Property 1 is trivially satisfied (by construction) in a single population, its generality to other populations is a hypothesis that needs to be tested empirically. The population against which others will be compared is designated as normed. The most stringent criterion for generality requires that the feature orders of all other populations match perfectly to that of the normed population. However, it is more practical to require only a high degree of matching in ordering. Accordingly, a measure that assesses the degree of matching is sought. We revisit this assessment issue later.

Property 2: Relative Consistency in Rating Central Features

In addition to being rated higher in prototypicality, some researchers argue that central features should also be rated more consistently than peripheral ones (Fehr & Russell, 1984; Mervis & Rosch, 1981). This consistency property can be reflected in the population standard deviations of the ratings. A stringent interpretation of this property is that $\sigma_1 < \sigma_2 < \cdots < \sigma_{35}$. However, because features are already ordered according to their mean prototypicality, it is unlikely that such a stringent interpretation of the consistency property would find any practical applications. Therefore, a more realistic interpretation is to view the consistency property only as a general trend of the ordered features. Suppose the ordered features are partitioned into two sets: a central set and a peripheral set. Let σ_c be the arithmetic mean of standard deviations of the central features and σ_p be the arithmetic mean of standard deviations of the peripheral features. A weaker consistency property is stated as follows.

In the population,

$$\sigma_c < \sigma_p$$

Once the consistency property is established in a normed population, the same property can become a criterion to gauge cross-cultural universality in other populations. Take, as an example, the 35 features of nostalgia identified by Hepper et al. (2012). The central feature set consists of the 18 most highly rated features. The peripheral feature set consists of the remaining 17 features. Property 2 requires that the average standard deviation (σ_c) of the 18 central features be smaller than the average standard deviation (σ_p) of the remaining 17 peripheral features.

Property 3: Distinctive Elevations of the Central and Peripheral Feature Sets

When there is no a priori reason to favor a particular partitioning scheme, splitting the ordered features into approximate halves is not an unreasonable initial step. To justify the interpretation of "central" and "peripheral" feature sets, however, a distinctiveness property of these feature sets is called for. Let μ_c be the mean rating of the central features and μ_p be the mean rating of the peripheral features. The following properties can be used to validate the distinction between the central and peripheral feature sets.

In the population,

$$\mu_c > \mu_p + \delta_1\sigma_p, \qquad \mu_p < \mu_c - \delta_2\sigma_c$$

Or, equivalently,

$$\frac{\mu_c - \mu_p}{\sigma_p} > \delta_1, \qquad \frac{\mu_p - \mu_c}{\sigma_c} < -\delta_2 \quad \left(\Leftrightarrow \frac{\mu_c - \mu_p}{\sigma_c} > \delta_2\right)$$

where δ_1 and δ_2 are distinctiveness criterion values. Given that the left sides of the above inequalities are standardized distances, it is useful to consider δ_1 and δ_2 as effect size measures (Cohen, 1988) for comparison purposes. A large effect size for distinguishing central and peripheral features is essential to prototype theory. Although a fixed number for defining a large effect size seems to be arbitrary, the guidelines provided by Cohen (1988) can serve as a good starting point. That is, in social science research, an effect size value of 0.8 is considered large, 0.5 is medium, and 0.2 is small. Therefore, in order to claim distinctiveness between the central and peripheral feature sets, the standardized distances must at least be larger than the medium effect size. This suggests that δ_1 or δ_2 must at least exceed 0.5 and ideally approximate 0.8. It is, then, reasonable to use the midpoint 0.65 as the criterion value for δ_1 or δ_2.

In discussing the two possible criteria for distinctiveness, we have not explicitly stated whether both or either one of the inequalities are required.[1] Whereas $\left(\mu_c - \mu_p\right)/\sigma_p$ is the standardized distance of the mean of *central* features from the distribution of *peripheral* features, $\left(\mu_p - \mu_c\right)/\sigma_c$ is the standardized distance of the mean of *peripheral* features from the distribution of *central* features. Although both standardized distances involve the difference between μ_c and μ_p in the numerators, their magnitudes are generally different due to standardizations via different distributions (in particular, via different standard deviations). Only when $\sigma_p = \sigma_c$ are the two inequalities equivalent, assuming that the criterion values δ_1 and δ_2 are the same. However, when σ_c and σ_p are different (and this is likely because in theory central features should be rated more consistently than peripheral features), three scenarios for the two inequalities are possible. In the first scenario, both inequalities are satisfied, and this is a clear-cut case to accept the distinctiveness of the adjacent feature sets in question. In the second scenario, both inequalities are not satisfied, and this is also a clear-cut case to reject distinctiveness. In the third scenario, one inequality is satisfied but the other is not. Should one accept or reject the distinctiveness in this case? We propose a combined criterion based on the average of the standardized distances. That is, the (combined) distinctiveness criterion (Property 3) requires that following inequality be satisfied:

$$\frac{1}{2}\left(\frac{\mu_c - \mu_p}{\sigma_p}\right) + \frac{1}{2}\left(\frac{\mu_c - \mu_p}{\sigma_c}\right) > \delta_3$$

As argued previously, $\delta_3 = 0.65$ is recommended for partitions with two feature sets. This combined criterion provides a simple, yet reasonable, quantitative way to determine distinctiveness when the two original inequalities are discordant. In fact, this combined criterion can be applied generally, because it is consistent with decisions on distinctiveness in the first two clear-cut

[1] We thank a reviewer for bringing this issue to our attention. If both equalities are required for distinctiveness, then the criterion can be simplified as

$$\frac{\mu_c - \mu_p}{\sigma_{max}} > \delta, \quad \text{where } \sigma_{max} = \max(\sigma_p, \sigma_c)$$

If only one of them is required for distinctiveness, then the criterion can be simplified as

$$\frac{\mu_c - \mu_p}{\sigma_{min}} > \delta, \quad \text{where } \sigma_{min} = \min(\sigma_p, \sigma_c)$$

However, to utilize more information for determining the distinctiveness of marginal cases, our proposal is based on the averaging of standardized distances. See text for explanations.

scenarios. That is, with all criterion values δ_1, δ_2, and δ_3 fixed at the same level, we observed the following:

1. When both inequalities basing on δ_1 and δ_2 are satisfied, the combined distinctiveness criterion basing on δ_3 would also be satisfied.
2. When both inequalities basing on δ_1 and δ_2 are *not* satisfied, the combined distinctiveness criterion basing on δ_3 would also *not* be satisfied.

For convenience, we apply the combined criterion in our analysis of distinctiveness.

Once the central/peripheral partitioning is justified by the distinctiveness property in a normed population, to assess cross-cultural universality, researchers can examine whether the same distinctiveness property holds in other populations of interest.

Depth of Partitioning

Stronger versions of the consistency (Property 2) and distinctiveness (Property 3) properties can be formulated upon further partitioning of the central and peripheral features. For example, Hepper et al. (2014) proposed four ordered partitioned sets of nostalgia features: C1$=\{a_1, a_2, \ldots, a_9\}$ (first nine central features), C2 $=\{a_{10}, a_{11}, \ldots, a_{18}\}$ (second nine central features), P1 $=\{a_{19}, a_{20}, \ldots, a_{27}\}$ (first nine peripheral features), and P2 $=\{a_{28}, a_{29}, \ldots, a_{35}\}$ (last eight peripheral features), respectively. The consistency property on these four partitioned sets (Property 2.1) is stated as:

In the population,

$$\sigma_{c1} < \sigma_{c2} < \sigma_{p1} < \sigma_{p2}$$

where the subscripts represent the feature sets.

The distinctiveness of these four partitioned sets (Property 3.1) can be validated by demonstrating the following properties in the population:

$$\text{(a)} \quad \mu_{c1} > \mu_{c2} + \gamma_{11}\sigma_{c2}\,,\ \ \mu_{c2} < \mu_{c1} - \gamma_{12}\sigma_{c1}\,,$$

$$\text{(b)} \quad \mu_{c2} > \mu_{p1} + \gamma_{21}\sigma_{p1}\,,\ \ \mu_{p1} < \mu_{c2} - \gamma_{22}\sigma_{c2}$$

$$\text{(c)} \quad \mu_{p1} > \mu_{p2} + \gamma_{31}\sigma_{p2}\,,\ \ \mu_{p2} < \mu_{p1} - \gamma_{32}\sigma_{p1}$$

where the subscripts C1, C2, P1, and P2 represent the feature sets and $\gamma_{11}, \gamma_{12}, \ldots$, and γ_{32} are distinctiveness criterion values. For four partitioned sets, the distinctiveness criterion values could be set to 0.35, corresponding to the cut-off between small and medium effect sizes (Cohen, 1988).

Similar to the development of a combined distinctiveness criterion for the case with two partitioned sets (Property 3), we rephrase three combined criteria for determining the distinctiveness of four partitioned sets (Property 3.1) as follows:

$$\text{(a)} \quad \frac{1}{2}\left(\frac{\mu_{c1}-\mu_{c2}}{\sigma_{c2}}\right)+\frac{1}{2}\left(\frac{\mu_{c1}-\mu_{c2}}{\sigma_{c1}}\right)>\gamma_{13}$$

$$\text{(b)} \quad \frac{1}{2}\left(\frac{\mu_{c2}-\mu_{p1}}{\sigma_{p1}}\right)+\frac{1}{2}\left(\frac{\mu_{c2}-\mu_{p1}}{\sigma_{c2}}\right)>\gamma_{23}$$

$$\text{(c)} \quad \frac{1}{2}\left(\frac{\mu_{p1}-\mu_{p2}}{\sigma_{p2}}\right)+\frac{1}{2}\left(\frac{\mu_{p1}-\mu_{p2}}{\sigma_{p1}}\right)>\gamma_{33}$$

Again, once the distinctiveness of the feature sets is established in a population, cross-cultural researchers can use this property as a criterion to assess cross-cultural universality in other populations.

Two comments about the criterion values are in order. First, the γs in Property 3.1 (four partitioned sets) should be smaller than the δs in Property 3 (two partitioned sets). Given that finer partitioning implies closer feature sets, a less stringent distinctiveness criterion is appropriate for finer partitioning. Hence, $\gamma < \delta$. Second, the γs in Property 3.1 can be of different importance. It may be relatively more important for Property 3.1b to hold (i.e., clear demarcation between central and peripheral features) than Property 3.1a or 3.1c (i.e., clear demarcation between the adjacent central or peripheral feature sets). A way to reflect the relative importance is to make criterion values in Property 3.1a or Property 3.1c smaller than those in Property 3.1b. For simplicity, we do not attempt this fine adjustment in the current chapter.

Even stronger versions of the consistency (Property 2) and distinctiveness (Property 3) properties can be stated by further partitioning of the four feature sets. Ultimately, continuing the partitioning process leads to the strongest prototype properties at the individual feature level. Depending on research domain, deeper partitioning might or might not be desirable. On the one hand, overly shallow partitioning, although easier for establishing cross-cultural universality, provides insufficient detail for adequate scientific understanding. On the other hand, overly deep partitioning might be too stringent and complicated to allow parsimonious interpretation. Therefore, some balance on depth of partitioning is needed. Moreover, it is possible that some constructs have uneven feature sets. For example, a construct can have one central feature set that consists of a single or a few strong features and a peripheral feature set that consists of many secondary features. In summary, the practicality and interpretability of the consistency and distinctiveness properties hinge on suitable depth of partitioning, which, in turn, depends on the interplay of subject domain, level of understanding being sought, and the state of knowledge about the construct in question. Hepper et al. (2014) expressed the consistency and distinctiveness properties as four partitioned sets of central and peripheral features of nostalgia (C1, C2, P1, and P2) with approximately equal sizes. For ease of exposition, we adopt their partitioning scheme.

Criteria for Cultural Universality

We proposed some structural properties of the features (or feature sets) of a complex construct under prototype theory. We now present criteria for establishing cross-cultural universality of a complex construct. Suppose individuals in another population or culture rate the same set of features. By using notation similar to those described previously but with a superscript (i) to identify this new population, the set of random variables for the feature set $A = \{a_1, a_2, a_3, \ldots\}$ are $x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, \ldots$. Similarly, notations for population means ($\mu^{(i)}$ s) and standard deviations ($\sigma^{(i)}$ s) are created for this new population.

Criterion 1: Similar Ordering of Features in New Populations

The prototypicality order of features in the feature set a_1, a_2, a_3, ...} in the new population should resemble the order in the normed population. Here, we propose a simple measure for the degree of resemblance in ordering. Let $\omega(\mu_j)$ and $\omega(\mu_j^{(i)})$ denote the rank orders of typicality of feature a_j in the normed and new populations, respectively. By construction, $\omega(\mu_j) = j$ for all j, but $\omega(\mu_j^{(i)}) = j$ is not necessarily true for each i (a given culture). By treating $\omega(\mu)$ and $\omega(\mu^{(i)})$ as vectors of ranks, the correlation between them is the rank correlation $\rho = \rho(\omega(\mu), \omega(\mu^{(i)}))$. Exact match in feature ordering is indicated when $\rho = 1$. Hence, the resemblance in feature ordering in two populations can be measured by ρ—the higher the more resemblance. We propose to require that ρ be greater than 0.7, to refine Criterion 1 as follows:

Criterion 1′: High rank-order correlations of features with the normed population.

Formally, the prototypicality order of features in feature set $A = \{a_1, a_2, a_3, \ldots\}$ in the new population should correlate at $\rho = \rho(\omega(\mu), \omega(\mu^{(i)})) = 0.7$ or higher to that of the normed population. Although 0.7 seems to be an arbitrary number, it becomes more interpretable when one looks at the corresponding requirement in ρ^2. Here it requires ρ^2 to be larger than 0.49, which means that at least half of the ranking variance in the new population must be explained by the ranking in the normed population.

Criterion 2: Relative Consistency in Rating Central Features in New Populations

The central features should be rated more consistently than the peripheral features in the new population i.[2] That is,

$$\text{(a)} \quad \sigma_{\text{c}}^{(i)} < \sigma_{\text{p}}^{(i)}$$

when the features are partitioned into the central and peripheral sets by using the same partitioning scheme in the normed population. The quantity $\sigma_{\text{c}}^{(i)}$ ($\sigma_{\text{p}}^{(i)}$) represents the arithmetic mean of standard deviations of the central (peripheral) features in population i. A stronger criterion with four partitioned sets is

$$\text{(b)} \quad \sigma_{\text{c1}}^{(i)} < \sigma_{\text{c2}}^{(i)} < \sigma_{\text{p1}}^{(i)} < \sigma_{\text{p2}}^{(i)}$$

Similarly, the quantities in the inequalities represent the arithmetic means of standard deviations of the features in sets C1, C2, P1, and P2, respectively, in population i.

Criterion 3: Distinctiveness of Feature Sets in New Populations

The partitioned sets of features should be distinct in the new population i, if they are distinct in the normed population under the same partitioning scheme. That is, for a two-level partitioning involving distinct central and peripheral feature sets, it requires

$$\text{(a)} \quad \mu_{\text{c}}^{(i)} > \mu_{\text{p}}^{(i)} + \delta_1 \sigma_{\text{p}}^{(i)}, \ \mu_{\text{p}}^{(i)} < \mu_{\text{c}}^{(i)} - \delta_2 \sigma_{\text{c}}^{(i)}$$

[2] Unlike the next two criteria, we stated the consistency criterion without considering effect sizes. The main reason is that a standardized scale for standard deviations, and hence the corresponding effect size measure, have not been well established.

where the criterion values δ_1 and δ_2 could be set at 0.65, corresponding to the midpoint between a medium (0.50) and large (0.80) effect size.

The following are stronger criteria for four distinct partitioned sets:

$$\text{(b)}\ \ \mu_{c1}^{(i)} > \mu_{c2}^{(i)} + \gamma_{11}\sigma_{c2}^{(i)} \text{ and } \mu_{c2}^{(i)} < \mu_{c1}^{(i)} - \gamma_{12}\sigma_{c1}^{(i)},$$

$$\text{(c)}\ \ \mu_{c2}^{(i)} > \mu_{p1}^{(i)} + \gamma_{21}\sigma_{p1}^{(i)} \text{ and } \mu_{p1}^{(i)} < \mu_{c2}^{(i)} - \gamma_{22}\sigma_{c2}^{(i)}, \text{ and}$$

$$\text{(d)}\ \ \mu_{p1}^{(i)} > \mu_{p2}^{(i)} + \gamma_{31}\sigma_{p2}^{(i)} \text{ and } \mu_{p2}^{(i)} < \mu_{p1}^{(i)} - \gamma_{32}\sigma_{p1}^{(i)}$$

where the criterion values γ could be set at 0.35, corresponding to an effect size that is intermediate between small (0.20) and medium (0.50). When a clear demarcation between central and peripheral features is more important than a clear demarcation between the adjacent central or peripheral feature sets, this can be reflected by adjusting the criterion values accordingly (i.e., setting γ_{21} and γ_{22} relatively higher).

Criterion 4: Similar Elevations of the Feature Sets in New Populations

Are Criteria 1–3 sufficient to establish cultural universality? Notice that all criteria established so far are more concerned with whether the "relative" structural properties (ordering, relative consistency in rating central features, and distinctiveness of feature sets) are preserved within new populations. Should the "absolute" elevations of features also be similar in new populations? We propose that they should, because the same rating scheme has been used for measuring prototypicality of features in each culture.

If the features are partitioned into central and peripheral sets, the following inequalities operationalize the elevation criterion:

$$\text{(a)}\ \ \left|\mu_c^{(i)} - \mu_c\right| < \beta_1\sigma_c \ \text{ and } \ \left|\mu_p^{(i)} - \mu_p\right| < \beta_2\sigma_p$$

where β_1 and β_2 are criterion values in terms of the standard deviations of the corresponding feature sets. If the features are partitioned into four feature sets, then the following inequalities operationalize the criterion:

$$\text{(b)}\ \ \left|\mu_{c1}^{(i)} - \mu_{c1}\right| < \beta_{11}\sigma_{c1} \ \text{ and } \ \left|\mu_{c2}^{(i)} - \mu_{c2}\right| < \beta_{12}\sigma_{c2}$$

$$\text{(c)}\ \ \left|\mu_{p1}^{(i)} - \mu_{p1}\right| < \beta_{21}\sigma_{p1} \ \text{ and } \ \left|\mu_{p2}^{(i)} - \mu_{p2}\right| < \beta_{22}\sigma_{p2}$$

where βs are criterion values.

When these βs approach zero, these criteria represent strict matching in elevations. Thus, a more reasonable requirement would be to set these criteria to a value that represents the upper boundary of a small effect size. Using the previous argument about effect size demarcation, 0.35 is chosen as the criterion value for βs (i.e., intermediate between a small and medium effect size).[3]

Summary

To establish cross-cultural universality of complex psychological constructs, one needs to show that its prototypical features are ordered with a high degree of similarity across cultures (Criterion 1), the central (vs. peripheral) features are more consistently rated by individuals across cultures (Criterion 2), the central and peripheral feature sets so partitioned are distinct across cultures (Criterion 3), and elevations of the feature sets should be similar across cultures (Criterion 4). Once these criteria are operationalized, researchers can derive the associated statistical analyses for samples. In the next section, we apply some conventional statistical tests to these criteria. We make no claim that these tests or analyses are optimal on statistical grounds, but, in the absence of unique tests for assessing these criteria, they allow researchers to use existing techniques to evaluate cross-cultural universality, thereby making such research questions more accessible.

Method and Results

In this section, we demonstrate, using data from Hepper et al. (2012) and Hepper et al. (2014), how the cross-cultural criteria we developed can be applied to the illustrative case of nostalgia conceptions in different cultures. Hepper et al. (2012) identified 35 central and peripheral features of nostalgia. Table 6.1 presents these features and their summary statistics in prototypicality rating. These results were based on a UK sample, which we designate as the normed UK sample hereafter. Given that Hepper et al.

[3] The validity of Criterion 4 assumes that biases due to translation and response styles in cultures are negligible. Given that it is usually difficult to distinguish such biases from true elevation differences, devising instruments that are culturally unbiased is of paramount importance.

(2012) validated these 35 features in several studies, we treat the normed UK sample as a reliable normed population (culture) for the purpose of assessing cross-cultural universality. We share the computer code for all analyses in Supplemental Materials.

The Structural Properties of Nostalgia Features in the Normed UK Sample

First, we should establish the structural properties of these 35 nostalgia features in the United Kingdom. Property 1 requires that features be ordered according to the prototypicalities. As shown in Table 6.1, in which the features are ordered by their mean prototypicality rating, Property 1 is trivially satisfied.

Table 6.1 also shows the average means and standard deviations of the partitioned feature sets. When the nostalgia features are partitioned into two sets, the average standard deviations are 1.54 and 1.82, respectively, for the central and peripheral sets. Hence, central features were rated more consistently, satisfying Property 2. When the nostalgia features are partitioned into four sets, the average standard deviations are 1.39, 1.69, 1.79, and 1.86, respectively, for the C1, C2, P1, and P2 feature sets. Hence, the consistency property still holds with the four partitioned sets (Property 2.1).

Property 3 requires that the partitioned sets are distinguishable. The average prototypicality ratings are 6.23 for the central set and 4.11 for the peripheral set. The standardized distances are 1.16 and −1.37, respectively, when using the central and peripheral feature sets as reference distribution. The average absolute standardized distance is 1.27. These distance measures clearly show a sizable separation (i.e., a large effect size) between the central and peripheral feature sets. With four sets, the corresponding averages are 6.60. 5.86, 4.81, and 3.34. The average absolute standardized distances are 0.49, 0.60, and 0.81, respectively, for comparing the C1/C2, C2/P1, and P1/P2 pairs. These distances represent at least medium effect sizes, which means that the four feature sets are still clearly distinguishable (Property 3.1).

Overall, the normed UK sample (i.e., the normed population that is used for studying cross-cultural universality) exhibits desirable structural properties under prototype theory. Next, we examine the cross-cultural universality of nostalgia conceptions by using the established criteria under a two-level and a four-level partitioning scheme.

Cross-Cultural Universality with Two-Level Partitioning: Central and Peripheral Features in New Populations

Criterion 1: Similar Ordering of Features

Hepper et al. (2014) asked participants ($N = 1{,}704$) to rate the relatedness (i.e., prototypicality) of the 35 nostalgia features identified by Hepper et al. (2012) across 18 countries. One of these countries was the United Kingdom and, hence, the 2014 study provides a means to assess the replicability of the 2012 results. We computed rank-order correlations by treating the 2012 UK sample as the normed population (Table 6.2). The countries are ordered by putting the 2014 UK sample first, followed by the other countries in descending order of their average central-feature rating.

Column 2, labeled "UK (Normed)," in Table 6.2 shows the rank correlations[4] of the nostalgia features in various countries with that of the normed 2012 UK sample. The first correlation, with the 2014 UK sample, is particularly high: 0.976. This confirms replicability. The remaining correlations in column 2 are all very high except for Cameroon, Poland, Romania, and Uganda, which have rank correlations lower than 0.7, although Poland and Romania are close to 0.7. Hence, it is safe to state that all countries, except Cameroon and Uganda, have similar prototypicality ordering of the nostalgia features. The third column in Table 6.2 shows the rank correlations between the 2014 UK sample and all other countries. Overall, the pattern of correlations is very similar to that observed in column 2. Therefore, in terms of feature ordering, we established that the nostalgia features are ordered similarly in most countries.

Criterion 2: Relative Consistency in Rating Central Features

Table 6.3 shows the average ratings (the "mean" column) and the average standard deviations of the features in the central and peripheral feature sets (the "SD" column) for the countries studied by Hepper et al. (2014), together with the normed UK sample of Hepper et al. (2012). In all countries, except for Uganda, the central (compared to peripheral) features were rated on average more consistently (i.e., with smaller standard deviations). Uganda has nearly identical standard deviations (consistency) in rating central and peripheral features.

[4] The ranks of features in countries are derived from the mean ratings of the features. Given that different sample sizes were used in different countries, the ranks and therefore the rank correlations in Table 6.2 have different degrees of reliability. The sample sizes range from 62 to 172 for these countries. See Table 6.3 for more details about sample sizes.

Table 6.2 Rank Correlations of Nostalgia Features of Various Countries with the United Kingdom

Country	UK (Normed)	UK
UK	0.976	
USA	0.948	0.948
Israel	0.927	0.927
Greece	0.857	0.866
China	0.798	0.822
Australia	0.968	0.960
Romania	<u>0.688</u>	<u>0.698</u>
Netherlands	0.851	0.843
Japan	0.909	0.906
Ireland	0.925	0.926
Turkey	0.822	0.849
Germany	0.856	0.870
Chile	0.822	0.827
India	0.889	0.898
Poland	<u>0.681</u>	0.718
Ethiopia	0.702	0.700
Cameroon	<u>0.643</u>	<u>0.665</u>
Uganda	<u>0.489</u>	<u>0.532</u>

Notes: The underlined entries are smaller than 0.7. The normed UK sample is from Hepper et al. (2012), and the sample for validation is from Hepper et al. (2014).

Criterion 3: Distinctiveness of Feature Sets

Table 6.3 further reveals that all countries have higher average ratings in the central than peripheral features. Statistical significance tests on the mean differences have been reported in Hepper et al. (2014, Table 4). All F tests for the mean differences were significant at the 0.0001 α level. The next three columns in Table 6.3 display the sample standardized distances between the central and peripheral feature sets. That is, $d1$ ($d2$) estimates how far away the average rating of central (peripheral) features is from the distribution of peripheral (central) features. We present the combined standardized distance $d3$, which is the average of $d1$ and $-d2$, in Table 6.3. This combined distance is to be compared with the criterion values δ_3 for determining distinctiveness. In the normed UK sample, the central features are not only rated higher on average than the peripheral features but also highly distinguishable from the peripheral features. The magnitudes of $d1$, $d2$, and $d3$ for the normed UK sample in Table 6.3 show that the standardized distances are all greater than 1, which

Table 6.3 Some Measures of the Central and Peripheral Nostalgia Features

	SD		Mean		Distinctiveness			*N*	Power
	C	P	C	P	*d*1	*d*2	*d*3		
UK (normed)	1.54	1.82	6.23	4.11	1.16	−1.37	1.27	102	0.99
UK	1.40	1.87	6.62	4.09	1.35	−1.81	1.58	97	0.99
USA	1.71	2.14	6.65	4.41	1.04	−1.31	1.18	165	0.99
Israel	1.61	2.06	6.41	3.93	1.20	−1.54	1.37	90	0.99
Greece	1.75	2.07	6.32	4.13	1.05	−1.25	1.15	172	0.99
China	1.69	1.99	6.24	4.40	0.93	−1.09	1.01	98	0.99
Australia	1.83	1.92	6.19	4.02	1.13	−1.18	1.15	66	0.99
Romania	1.84	2.17	6.12	4.57	0.71	−0.84	0.78	80	0.99
Netherlands	1.48	1.68	6.06	3.95	1.25	−1.43	1.34	120	0.99
Japan	1.72	2.00	6.00	4.44	0.78	−0.91	0.85	96	0.99
Ireland	1.82	1.88	5.93	4.78	0.61	−0.63	0.62	85	0.99
Turkey	2.03	2.28	5.89	3.74	0.94	−1.05	1.00	79	0.99
Germany	1.66	1.81	5.87	3.53	1.29	−1.41	1.35	84	0.99
Chile	1.79	2.01	5.78	3.78	0.99	−1.12	1.06	82	0.99
India	1.77	1.90	5.73	4.51	0.65	−0.69	0.67	68	0.99
Poland	1.76	1.95	5.69	3.88	0.93	−1.03	0.98	70	0.99
Ethiopia	2.17	2.34	5.56	4.46	0.47	−0.51	0.49	62	0.99
Cameroon	2.55	2.64	5.27	4.10	0.45	−0.46	0.45	102	0.99
Uganda	**1.84**	**1.84**	4.71	3.85	0.47	−0.47	0.47	88	0.99

Notes: The underlined entries for *SD*s are not showing the increasing pattern. The underlined entries for *d*3 are values that are not larger than the distinctive criterion value $\delta = 0.65$. Entries for power are computed using $\alpha = 0.05$ for testing a null hypothesis of no effect given the sample sizes of the countries and a true effect size of 0.65.

indicates a sizable separation between the central and peripheral feature sets. Indeed, for most of the countries studied, Table 6.3 illustrates that they do have acceptable or high degrees of distinctiveness between central and peripheral features, as indicated by *d*3 values that are at least as large as the criterion value, $\delta = 0.65$, in 14 out of 18 countries. Only Ireland, Cameroon, Ethiopia, and Uganda have *d*3 smaller than 0.65.

The last two columns of Table 6.3 display, respectively, the sample sizes and the power of rejecting the null hypothesis of no effect at 0.05 α level given the sample sizes of the countries and a true effect size of 0.65. The high power values (0.99 for all) indicate that the statistical tests have high sensitivity of detecting such a specified effect size, and they have good protection against the Type II error. That is, the high sensitivity of effect detection ensures that the distinctiveness in feature sets would be detected reliably, if they were present; and the good protection against the Type II error means that non-distinctiveness in feature sets, if concluded from the hypothesis tests, would be unlikely to result from sampling errors. Certainly, the observed high power values are in part due to the relatively large sample sizes, implying that the *d*1, *d*2, and *d*3 values are precise estimates of population standardized differences for carrying out trustable hypothesis tests.

Criterion 4: Similar Elevations of the Feature Sets

We turned next to absolute elevations of central and peripheral features. Figure 6.1 depicts the mean ratings of such feature sets with their 95% confidence intervals. Countries are ordered by their average central-feature rating after the 2014 UK sample. To set up acceptance regions around the elevations in the normed 2012 UK sample, we used the proposed $\beta = 0.35$ criterion. That is, if a country has a mean rating of a given feature set (i.e., central or peripheral) that is within 0.35 standardized distance of the same feature set in the normed 2012 UK sample, it is accepted as having a similar elevation (i.e., a small departure in terms of effect size). Hence, the vertical bars in Figure 6.1 mark the acceptance regions of the central and peripheral feature sets. A country that has an entire 95% confidence interval located within its corresponding acceptance region demonstrates the strongest evidence of similar elevations as that of the normed 2012 UK sample. For example, in Figure 6.1, all countries starting from the United Kingdom down to Japan (except for the United States, which has a slightly elevated central feature set) demonstrate the strongest evidence of similar central and peripheral elevations. On the contrary, strong evidence against similar elevation is displayed if the entire 95% confidence interval falls outside of its corresponding acceptance region. For example, Cameroon and Uganda have unacceptable (with regard to Criterion 4) low elevation of the central features. Whereas Poland and Ethiopia have marginally acceptable[5] elevation in the central features, Ireland has a marginally acceptable elevation in the peripheral features. Overall, Figure 6.1 shows that for most countries, central and peripheral features are elevated at similar levels to those of the normed 2012 UK sample.[6]

So far, the cultural universality of nostalgia conceptions is supported in most countries based on their similarity in feature ordering (Criterion 1),

[5] This means that less than half of the confidence interval of the average rating of a feature type overlaps with the acceptance region.

[6] We caution about the validity of the type of comparisons shown in Figure 6.1 in small samples. In applications, like the current study, the width of confidence intervals should be smaller than the acceptance regions. Otherwise, the confidence intervals can largely overlap with the acceptance regions simply due to large sampling errors. In general, researchers can increase the sample size to ensure that the confidence intervals are narrow enough for meaningful comparisons with the acceptance regions. Note that this was not an issue in the current data set.

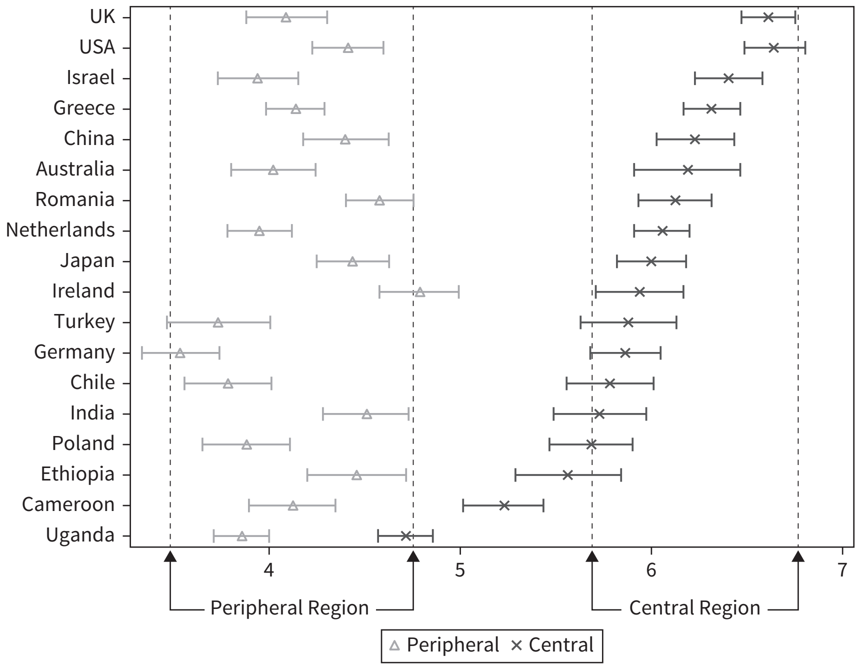

Figure 6.1 Confidence intervals of average ratings of the central and peripheral nostalgia features. Dashed vertical lines indicate the acceptance regions around the normed UK sample (criterion $\beta = 0.35$).

more consistent ratings of central than peripheral features (Criterion 2), distinctiveness of the central and peripheral features (Criterion 3), and similar elevations in central and peripheral features (Criterion 4). Uganda, Cameroon, and perhaps Ethiopia cast the most doubt because their central and peripheral features are not as distinct as in the normed 2012 UK sample (although still distinctive at a less stringent level). Also, their elevations of the central features are much lower than the expected level.

Cross-Cultural Universality with Four-Level Partitioning: C1, C2, P1, and P2 in New Populations

We now repeat the same analyses but with the four partitioned feature sets, C1 (first nine features), C2 (second nine features), P1 (third nine features), and P2 (last eight features). Given the finer level of analysis, the results will almost certainly render some countries in doubt for establishing cross-cultural universality. On the positive side, however, we may be able to draw stronger conclusions and have finer interpretations of the results. Note that the assessment of Criterion 1 (i.e., similar ordering of features) by means of rank-order correlations is not affected by depth of partitioning.

Criterion 2: Relative Consistency in Rating Central Features

Table 6.4 portrays the average standard deviations ("SD" columns) and means ("Mean" columns) of the features in feature sets, and Figure 6.2 graphs the numerical values. The *SD* columns in Table 6.4 display the average standard deviations of the features in C1, C2, P1, and P2 feature sets. The majority of the countries (11 out of 18) show an increasing trend in *SD*s, confirming the consistency property. Several countries do not exhibit the increasing *SD* trend in some adjacent pairs of feature sets. These discordant pairs are underlined in Table 6.4. Although these marked pairs do not seem to have a systematic pattern, the C1 features were rated more consistently than other features in all countries.

Table 6.4 Average Means and Standard Deviations of the Four Nostalgia Feature Sets

	SD				Mean			
	C1	C2	P1	P2	C1	C2	P1	P2
UK (normed)	1.39	1.69	1.79	1.86	6.60	5.86	4.81	3.34
UK	1.08	1.72	1.83	1.91	7.03	6.21	4.80	3.29
USA	1.59	1.82	2.11	2.17	6.85	6.44	4.92	3.84
Israel	1.44	1.77	2.11	2.00	6.74	6.08	4.51	3.28
Greece	1.59	1.90	2.04	2.11	6.63	6.01	4.76	3.43
China	1.62	1.76	1.97	2.01	6.35	6.13	4.64	4.13
Australia	1.70	1.96	1.99	1.85	6.56	5.81	4.40	3.59
Romania	1.69	1.99	2.11	2.24	6.43	5.82	4.66	4.48
Netherlands	1.32	1.63	1.72	1.63	6.47	5.65	4.41	3.42
Japan	1.66	1.78	1.97	2.03	6.19	5.82	4.68	4.16
Ireland	1.75	1.89	1.86	1.91	6.20	5.67	4.92	4.63
Turkey	2.02	2.05	2.33	2.23	6.02	5.75	3.98	3.47
Germany	1.52	1.81	1.89	1.73	6.15	5.60	3.84	3.18
Chile	1.77	1.82	1.97	2.06	5.90	5.66	3.96	3.57
India	1.75	1.80	1.87	1.93	5.94	5.53	4.77	4.21
Poland	1.62	**1.90**	**1.87**	2.04	5.98	5.39	3.89	3.86
Ethiopia	2.17	2.18	2.33	2.35	5.66	5.46	4.80	4.07
Cameroon	2.54	2.56	2.66	2.61	5.12	5.43	4.27	3.90
Uganda	1.80	1.88	1.83	1.85	5.05	4.37	3.91	3.79

Notes: The underlined entries for *SD*s are not showing the increasing pattern. The underlined entries for means are not showing the decreasing pattern.

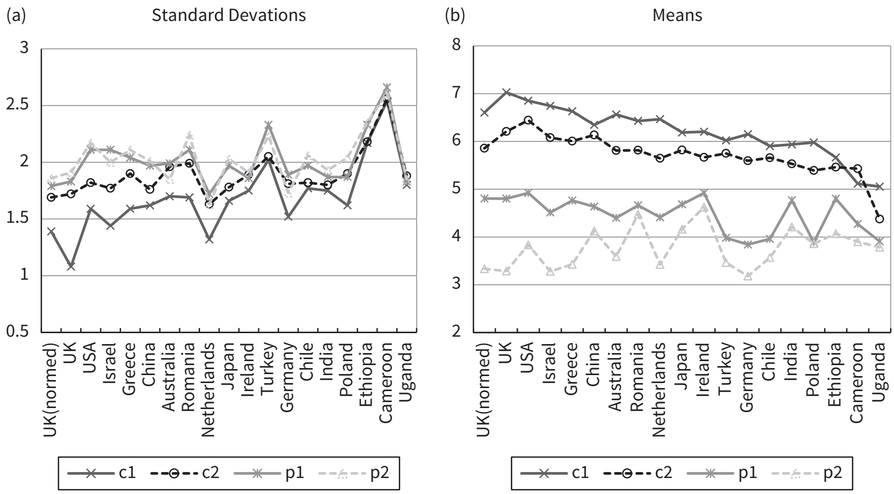

Figure 6.2 Average standard deviations and average mean ratings of C1, C2, P1, and P2 feature sets.

Figure 6.2a shows the trends of the *SD* values of the feature sets. The C1 features have the lowest *SD* values in all countries. The *SD* values of C2, P1, and P2 features in most countries (except for the United Kingdom, the United States, Israel, Greece, and China) do not evince clear patterns. Therefore, the consistency in rating central features seems to hold only when one compares the C1 features with other feature sets.

Criterion 3: Distinctiveness of Feature Sets

Table 6.4 shows that with the exception of Cameroon, all countries have a clear monotonic decreasing ordering of the average C1, C2, P1, and P2 ratings. For Cameroon, C2 had a higher average rating than C1. Figure 6.2b demonstrates the same pattern, but it depicts some useful trends. First, the C2/P1 separation is clear in all countries. This echoes the distinctiveness of the central and peripheral features reported in the preceding section. Second, when the average central feature rating decreases, the average ratings of the four feature sets become more similar (or the feature sets become less distinctive).

To address statistically the distinctiveness of the C1, C2, P1, and P2 feature sets, one can first test the mean differences in Table 6.4 by analysis of variance tests. Hepper et al. (2014, Table 3) conducted *F* tests for comparing adjacent partitioned sets and found that most mean differences were statistically significant. The only nonsignificant mean differences (at $\alpha = 0.05$) pertained to the P1/P2 comparisons in Poland, Romania, and Uganda and the C1/C2 comparison in Ethiopia (the C1/C2 reversal in Cameroon was significant).

However, the distinctiveness property of these feature sets requires more than a significant mean difference from 0. Table 6.5 shows the standardized distances between the adjacent feature sets. For the 2012 UK (normed) and 2014 UK samples, all *g*3 values are larger than the criterion value $\gamma = 0.35$, which marks the cutoff between small and medium effect sizes. Hence, the distinctiveness property is clear for all feature sets in the United Kingdom. For other countries, the C2/P1 distinctiveness ($g3 > 0.35$) is strongly supported in 15 out of 17 countries. The C2/P1 distinctiveness in Ethiopia and Uganda is doubtful. For C1/C2 and P1/P2 distinctiveness, the support is weaker. Only 5 out of 17 countries support the distinctiveness between C1 and C2, and only 6 countries support the distinctiveness of P1 and P2.

Table 6.5 Distinctiveness of the Four Partitioned Sets

	C1 vs. C2			C2 vs. P1			P1 vs. P2			N	Power
	*g*1	*g*2	*g*3	*g*1	*g*2	*g*3	*g*1	*g*2	*g*3		
UK (normed)	0.41	−0.54	0.49	0.59	−0.62	0.60	0.79	−0.82	0.81	102	0.93
UK	0.48	−0.76	0.62	0.77	−0.82	0.79	0.79	−0.83	0.81	97	0.93
USA	0.22	−0.26	<u>0.24</u>	0.72	−0.84	0.78	0.50	−0.51	0.51	165	0.99
Israel	0.38	−0.46	0.42	0.74	−0.89	0.81	0.62	−0.59	0.60	90	0.91
Greece	0.33	−0.39	0.36	0.61	−0.66	0.63	0.63	−0.65	0.64	172	0.99
China	0.12	−0.13	<u>0.13</u>	0.76	−0.85	0.81	0.25	−0.26	<u>0.26</u>	98	0.93
Australia	0.38	−0.44	0.41	0.71	−0.72	0.71	0.44	−0.41	0.42	66	0.80
Romania	0.31	−0.36	<u>0.34</u>	0.55	−0.58	0.56	0.08	−0.09	<u>0.09</u>	80	0.87
Netherlands	0.50	−0.62	0.56	0.72	−0.76	0.74	0.60	−0.57	0.59	120	0.97
Japan	0.21	−0.22	<u>0.21</u>	0.58	−0.64	0.61	0.26	−0.26	<u>0.26</u>	96	0.92
Ireland	0.28	−0.31	<u>0.30</u>	0.40	−0.39	0.40	0.15	−0.16	<u>0.16</u>	85	0.89
Turkey	0.13	−0.13	<u>0.13</u>	0.76	−0.86	0.81	0.23	−0.22	<u>0.23</u>	79	0.87
Germany	0.31	−0.37	<u>0.34</u>	0.93	−0.97	0.95	0.38	−0.35	0.37	84	0.89
Chile	0.13	−0.14	<u>0.14</u>	0.86	−0.94	0.90	0.19	−0.20	<u>0.19</u>	82	0.88
India	0.23	−0.23	<u>0.23</u>	0.41	−0.43	0.42	0.29	−0.30	<u>0.29</u>	68	0.81
Poland	0.31	−0.36	<u>0.34</u>	0.80	−0.79	0.80	0.01	−0.02	<u>0.02</u>	70	0.82
Ethiopia	0.09	−0.09	<u>0.09</u>	0.28	−0.30	<u>0.29</u>	0.31	−0.31	<u>0.31</u>	62	0.77
Cameroon	−0.12	0.12	<u>−0.12</u>	0.43	−0.45	0.44	0.14	−0.14	<u>0.14</u>	102	0.94
Uganda	0.36	−0.38	0.37	0.26	−0.25	<u>0.25</u>	0.07	−0.07	<u>0.07</u>	88	0.90

Notes: *g*1 indicates how many average standard deviations the mean of the first feature set is away from the distribution of the second feature set. *g*2 indicates how many average standard deviations the mean of the second feature set is away from the distribution of the first feature set. *g*3 is the average of the values of *g*1 and −*g*2. The underlined entries for *g*3 are values that are not larger than the distinctiveness criterion value $\gamma = 0.35$. Entries for power are computed using $\alpha = 0.05$ for testing a null hypothesis of no effect given the sample sizes of the countries and a true effect size of 0.35.

The last two columns of Table 6.5 display, respectively, the sample sizes and the power of rejecting the null hypothesis of no effect at 0.05 α level given the sample sizes of the countries and a true effect size of 0.35. Except for Ethiopia, which had power of 0.77, the power for other countries is at least 0.8. As in Table 6.3, these high power values indicate that the statistical tests have great sensitivity of detecting distinctiveness in feature sets given the specific distinctiveness criterion and have good protection against the Type II error (i.e., against false claims of non-distinctiveness in feature sets). Again, the observed high power values are in part due to the relatively large sample sizes, implying that the *g*1, *g*2, and *g*3 values are precise estimates of population standardized distances for carrying out trustable hypothesis tests. Earlier we proposed that C1/C2 or P1/P2 distinction might not be as important so that a lower criterion for *g*3 could be used. For example, if *g*3 is set to 0.2, 12 out of 17 countries would support the C1/C2 distinctiveness, and 11 would support the P1/P2 distinctiveness. Even so, C1/C2 distinctions are in doubt for China, Turkey, Chile, Ethiopia, and Cameroon (note also that C2 is higher than C1 in Cameroon). The P1/P2 distinctions for Romania, Ireland, Chile, Poland, Cameroon, and Uganda are also in doubt.

Criterion 4: Similar Elevations of the Feature Sets

Next, we turned to the absolute elevations of the four feature sets. Figure 6.3 depicts the mean ratings of the four feature sets with 95% confidence intervals. Originally, we constructed this figure in the same way as Figure 6.1 with the use of the 0.35 criterion for setting up acceptance regions. However, the acceptance regions tended to overlap, making the acceptance criteria ambiguous. We therefore shrank the acceptance regions by using a stricter criterion (i.e., smaller value). Figure 6.3 uses the 0.24 criterion that results in "just" non-overlapping regions. Turkey, Chile, India, Poland, and the three African countries present strong evidence against similar C1 levels: All have much lower C1 elevations than the UK norm. Particularly questionable are the C1 elevations of Cameroon and Uganda because they are at the normed P1 level. The C2 elevations of all countries, except Uganda, overlap with the acceptance regions. Again, Uganda has a very low C2 elevation. The United States has C2 elevation that is as high as that of the C1 features. Lower than expected P1 elevations are observed in Germany, Chile, Poland, Turkey, and Uganda. Higher than expected P2 elevations are observed in China, India, Ireland, Japan, and Romania. These latter countries emphasize strongly the most peripheral features of nostalgia.

The analysis based on four feature sets provides more information about the universality of nostalgia conceptions than the one based on two feature

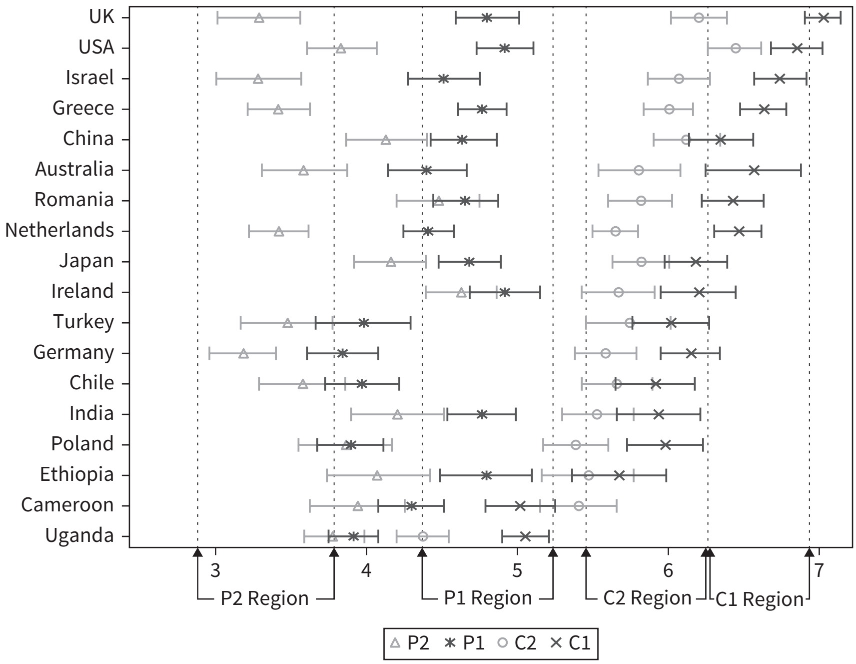

Figure 6.3 Confidence intervals of average ratings of the C1, C2, P1, and P2 nostalgia features. Dashed vertical lines indicate the acceptance regions around the normed UK sample (criterion $\beta = 0.24$).

sets (i.e., central vs. peripheral). First, it reveals that the consistency of rating central features (Criterion 2) occurs mainly in the C1 feature sets (the first nine features). Despite some irregularities, the C1 features were always rated more consistently. Second, it confirms a stronger ordering property in the ratings of central and peripheral feature sets (Criterion 3: C1 > C2 > P1 > P2 is observed in all countries but Cameroon, which has a mild violation in that C2 > C1). Although the distinctiveness between central and peripheral features is confirmed by comparing the elevations of C2 and P1, some countries do not have distinct C1/C2 (within central) or P1/P2 (within peripheral) feature sets. These irregularities occurred in five and six countries, respectively, for the C1/C2 and P1/P2 comparisons. Finally, the absolute elevations of the four feature sets (Criterion 4) clarify our interpretations of the irregularities observed in the central/peripheral features (see Figure 6.1). For example, the three African countries have lower elevations for the central features for different reasons. As Figure 6.3 shows, whereas Uganda is extremely low in both C1 and C2, Cameroon and Ethiopia are both low in C1 only. Although the C2 levels of the latter countries still overlap with the acceptance regions, their C1 and C2 features themselves also overlap. As another example, Ireland and Romania have elevated peripheral features in Figure 6.1. Figure 6.3 illustrates that this is due to elevated ratings of the P2 features, making them more on par with P1 features. This pattern also presents itself in the analysis of P1/P2 distinctiveness of these two countries in Table 6.5. Perhaps nostalgia has a more negative meaning for Ireland and Romania than for the United Kingdom and other countries. Poland, Cameroon, and Uganda have overlapping P1 and P2 as well, but they overlap at the middle of the normed P1 and P2 regions.

Conclusions from Confirmatory Analysis of Cross-Cultural Universality

The overall conclusion from the preceding analyses is that except for the African countries, nostalgia conceptions are, by and large, cross-culturally universal in terms of similar rank-ordering of the nostalgia features (Criterion 1), relative consistency in rating more central features (Criterion 2; especially for C1, but with minor irregularities in other feature sets in five countries), high distinctiveness of the central and peripheral features (Criterion 3; although with a small number of countries showing ambiguous C1/C2 or P1/P2 distinctions upon a finer four-level partitioning of the features), and high degree of agreement in absolute rating levels of central and peripheral features (Criterion 4; although with some countries showing lower C1, lower P1, or higher P2 upon a finer partitioning of the features). The nostalgia conceptions of Romania and Ireland emphasize more negative features of nostalgia compared with the United Kingdom. Poland and Uganda might have indistinguishable P1 and P2 features.

The three African countries evince similar prototypical orderings of the nostalgia features as the United Kingdom. However, their prototypicality ratings of central and peripheral features are much closer and, hence, less distinguishable. Perhaps one might still claim a weak cross-cultural universality for these African countries based on the similar ordering of features (Criterion 1) alone. However, one must also acknowledge that each of these countries might have some unique conceptions of the construct of "nostalgia," distinguishing them from the other 15 countries and from each other. Hepper et al. (2014) attempted to identify these potentially unique conceptions by inviting participants to list features that were not captured by the list provided. However, few participants listed additional features, and there was no evidence that particular additional features were listed only in some cultures.

The case of these African countries highlights how our proposed techniques can help identify where further targeted research is still needed.

Exploratory Techniques to Identify Homogeneous Clusters of Countries

Although the results in the preceding section support the cultural universality of nostalgia conceptions in most countries, some countries are identified to have mean rating patterns that deviate from the normed UK sample. For example, as depicted in Figures 6.1 and 6.3, African countries have significantly lower ratings of central (or C1) features, indicated by associated confidence intervals that do not overlap with the specified acceptance regions. To examine more systematically possible homogeneous groups of countries based on mean patterns, we used multivariate statistical techniques, such as cluster analysis and multidimensional scaling (MDS).

Hepper et al. (2014) presented a cluster analysis of the countries by using the mean ratings of the 35 features. They identified four clusters (groups) of countries. To further understand and interpret the mean patterns of these four clusters, we conduct an MDS analysis of the 35 features among the 18 countries. Before so doing, however, we first recapitulate the cluster analysis results of Hepper et al. (2014). Figure 6.4 shows the dendrogram from this analysis. The authors adopted a four-cluster solution. Reading from the bottom of the vertical axis, the first cluster includes the United Kingdom, the United States, Greece, Israel, the Netherlands, and Australia. This group has mean patterns that are most similar to the normed UK sample. Incidentally, these countries are located at the top of the chart in Figure 6.3, representing countries that have high C1 ratings. The next cluster that is closest to the first includes Romania, Ireland, India, Ethiopia, Japan, and China. Most countries in this cluster are located in the middle of the chart in Figure 6.3. They have medium C1 ratings. The third cluster includes Uganda and Cameroon, which are located at the bottom of the chart in Figure 6.3. These two countries have the lowest C1 ratings. The last cluster includes Poland, Germany, Turkey, and Chile. These countries also have medium C1 ratings, albeit somewhat lower than those in the second cluster.

We have associated the clusters with the C1 ratings in Figure 6.3. This is an initial interpretation of how these clusters might differ. Next, we demonstrate how MDS can offer more refined interpretations. To conduct MDS, a distance measure has to be used for quantifying the similarity between countries. In this regard, Hepper et al. (2014) used the absolute difference in mean ratings

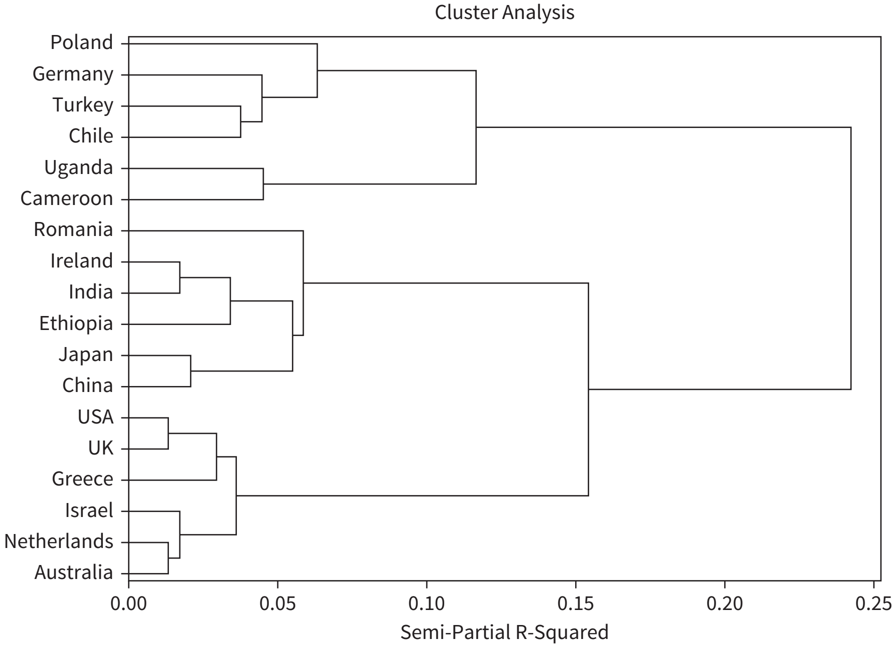

Figure 6.4 Dendrogram for a cluster analysis based on mean ratings of the nostalgia feature sets (adapted from Hepper et al., 2014).

of each feature to assess similarity. Hence, we used 35 matrices of similarity measures (for 35 features) for the 18 countries as input for the MDS analysis. The goal was to find the coordinates (or locations) of the countries in a multidimensional space that would give a satisfactory account of the observed feature similarities in those 35 matrices.

Figure 6.5 depicts the countries in two-dimensional space according to the MDS results. The four ovals in Figure 6.5 demarcate the four clusters of countries identified by Hepper et al. (2014). This replication of the four clusters provides a sound foundation for using the two-dimensional MDS solution to interpret the corresponding cluster results—there would be no need to resort to a higher dimensional MDS solution. The most common way to interpret an MDS result is to hypothesize the underlying latent dimensions by inspecting the objects (countries) in the multidimensional space. For example, for Dimension 1, the United States and United Kingdom are at the lower end, and Cameroon and Uganda are at the upper end. This could suggest that Dimension 1 reflects Westernization. However, the position of some other countries on Dimension 1 is inconsistent with this interpretation. For example, China is closer to the lower end, and Germany is closer to the upper end. Therefore, we propose an alternative strategy that is more objective and descriptive.

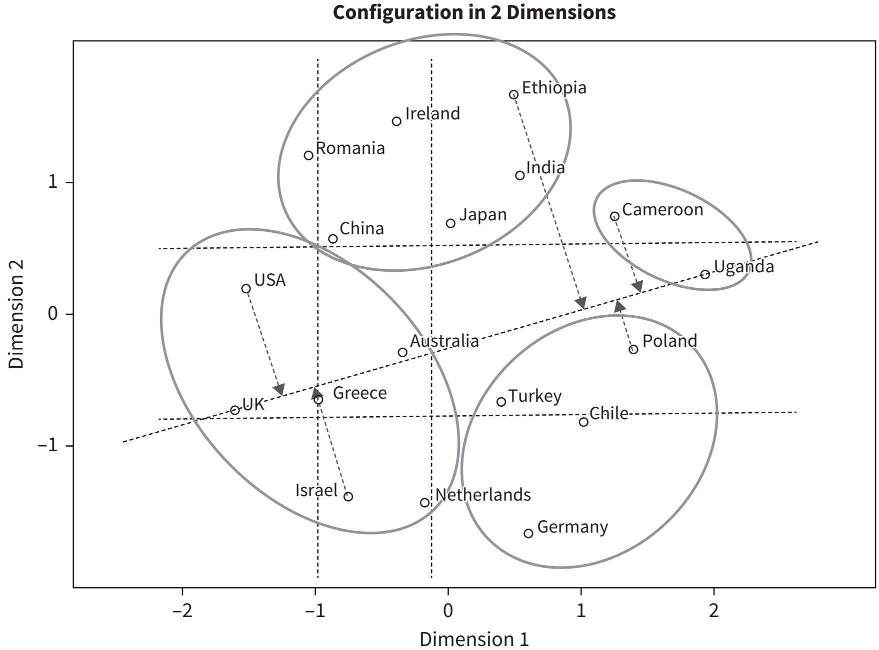

Figure 6.5 Two-dimensional representation of the multidimensional scaling results.

To interpret Dimension 1, we draw two horizontal lines in Figure 6.5 so that each captures a handful of countries that are approximately at the same level of Dimension 2. By so doing, we attempt to isolate the interpretation of Dimension 1 from Dimension 2. The upper horizontal line in Figure 6.5 connects approximately the United States, China, Japan, Cameroon, and Uganda. The lower horizontal line in Figure 6.5 connects approximately the United Kingdom, Greece, Turkey, and Chile. The left panel of Figure 6.6 shows the elevations of the C1, C2, P1, and P2 feature sets for the first group of countries. The right panel shows the elevations of these feature sets for the second group. The two panels show a common pattern, such that the curves in both plots converge as they move from left to right. Given that only the P2 curve is relatively flat in these two plots, the "reason" for such convergence is the declining trends in the C1, C2, and P2 curves. Hence, Dimension 1 in Figure 6.5 can be characterized as a general declining prototypicality of the C1, C2, and P1 features, resulting in reduced distinctiveness of the feature sets along the dimension. For example, Uganda, Cameroon, Poland, and Chile are high on Dimension 1, and all have relatively low elevations in the C1, C2, and P1 features.

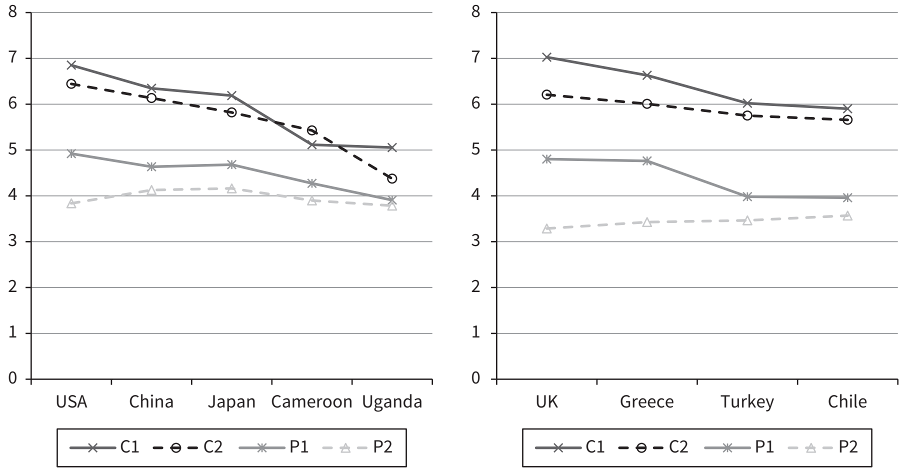

Figure 6.6 Interpretation of Dimension 1 of the mean patterns based on MDS analysis. In each panel, the four countries differ only on Dimension 1 (countries in the left panel are high on Dimension 2 and those in the right panel low on Dimension 2).

Similarly, to interpret Dimension 2, we draw two vertical lines in Figure 6.5 so that each line captures countries that differ only in Dimension 2, creating two groups of countries. The elevations of the feature sets for the countries along the left vertical line are plotted in the left panel of Figure 6.7. The right panel of Figure 6.7 plots the elevations for the countries along the right vertical line. Like those in Figure 6.6, these two plots show some convergence of the curves. Unlike those in Figure 6.6, however, the C1 and C2 elevations in Figure 6.7 do not evince strong decreasing patterns. These curves stay approximately at the same level along the dimension. A commonality of these two plots is that the P2 curve depicts a strong increasing trend. Therefore, Dimension 2 in Figure 6.5 indicates mainly an increasing emphasis of peripheral features (including negative features) for representing nostalgia. For example, Romania, Ireland, Ethiopia, and India are high on Dimension 2.

Given that the nostalgia features have been established in the United Kingdom, it would be interesting to look for a single indicator that can assess the similarity of each country to the United Kingdom in the MDS solution. Given that the United Kingdom is located at the extreme southwest end in Figure 6.5, one can start by drawing a line that connects the United Kingdom and a country in the farthest northeast direction to indicate a derived dimension in Figure 6.5. Hence, we drew a line between the United Kingdom and Uganda in Figure 6.5 to indicate such a derived dimension. Essentially, the dissimilarity of each country to the United Kingdom is indicated by the distance of its projection on the derived dimension to the United Kingdom. For example, Figure 6.5 shows projections of the United States, Israel, Ethiopia,

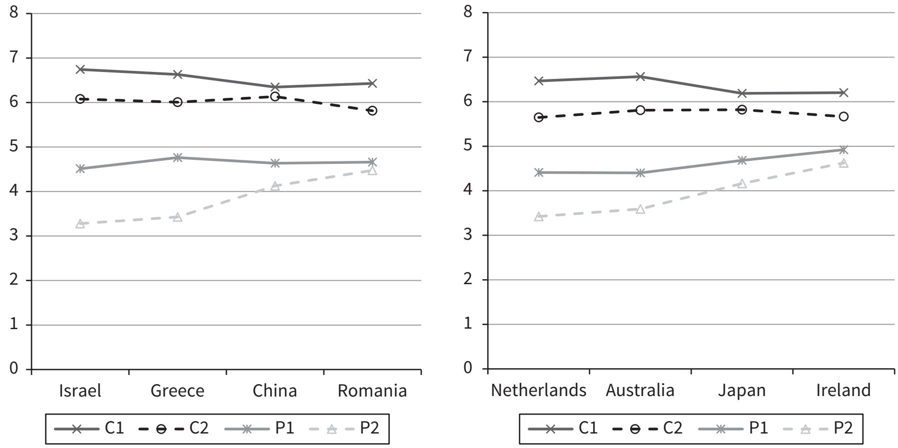

Figure 6.7 Interpretation of Dimension 2 of the mean patterns based on the MDS analysis. In each panel, the four countries differ only on Dimension 2 (countries in the left panel are high on Dimension 1 and those in the right panel are low on Dimension 1).

Poland, and Cameroon on the derived dimension. The United States is most similar to the United Kingdom, followed by Israel and Greece. At the other end, Uganda is the most dissimilar, followed by Cameroon, Poland, and Ethiopia. Finally, because this derived dimension is a combination of Dimensions 1 and 2 (leaning toward Dimension 1 more), countries that are captured by the lines that run parallel to the derived dimension should "combine" the trends of Dimensions 1 and 2. Figure 6.8 plots the C1, C2, P1, and

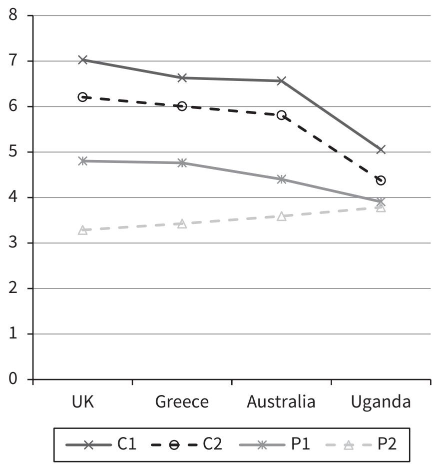

Figure 6.8 Interpretation of the derived dimension for characterization of the mean patterns.

P2 curves for four countries (the United Kingdom, Greece, Australia, and Uganda) that are located approximately along the derived dimension. Indeed, the C1, C2, and P1 curves are decreasing (i.e., the Dimension 1 characteristic) and the P2 curve is increasing (i.e., the Dimension 2 characteristic) along the derived dimension, resulting in less distinctive feature sets at the lower end of the dimension. Thus, the MDS analyses allow us to identify the key dimensions that delineate groups of countries as well as express numerically the countries that are most similar or different from a normed population.

Comparison with the Confirmatory Factor Analysis Approach

In the past few decades, CFA techniques have become popular in analyzing cross-cultural data (Matsumoto & Van de Vijver, 2011; see also Byrne et al., 1989; Byrne & Watkins, 2003). The CFA approach fits multiple-group models (Jöreskog, 1971) using structural equation modeling software such as EQS (Bentler, 2006), LISREL (Jöreskog & Sörbom, 1996), MPlus (Muthén & Muthén, 2012), or PROC CALIS of SAS/STAT (SAS Institute, 2014). Under the CFA framework, cultural equivalence or universality of constructs correspond to specific sets of invariant (or equality-constrained) parameters across cultures in a multiple-group CFA. Overall equivalence is supported if the invariance model satisfies some agreed-upon model-fit criteria. If the overall equivalence is unsupported, partial-invariance models that fit the data are searched manually with the aid of post hoc analytic tools such as Lagrange multiplier (LM) tests. As a byproduct of the search process, noninvariant items are detected to explain cross-cultural differences.

Due to the repeated fitting and refitting process for finding a good model for the data, conducting a multiple-group CFA (especially when there are more than a few groups/countries) could be a problematic and tedious process. To illustrate such a process, we applied the CFA approach to the current data. We report data-analytic details and results in Supplemental Materials. Here, we summarize three main analytic stages and the corresponding results.

In the first stage, we conducted exploratory factor analyses using the combined UK samples to establish a reasonable confirmatory factor pattern for subsequent multiple-group CFAs. After fitting models with three to six factors, we selected a four-factor solution because it accounted for 82% of common variance and its factor pattern was the most interpretable as well as compatible with prototype theory. We present the final rotated factor pattern with four factors in Table 6.6. We do not show factor loadings lower than 0.3, and we permuted the factor columns to improve interpretation of the factors. The parenthesized values after the features indicate their prototypicality order in the normed UK sample. Factors 1 and 4 are clearly identified with, respectively, the most central (C1) and most peripheral features (P2) of nostalgia. However, it is less clear which of Factors 2 or 3 is more central or peripheral. Furthermore, the five loadings that are *in light shades* are not considered indicative of the corresponding factors because the associated variables have larger loadings on other factors. Accordingly, we specified an initial confirmatory factor pattern with a simple structure by using only the remaining loadings shown in Table 6.6.

In the second stage, we modified repeatedly the initial confirmatory factor pattern with the goal of obtaining a final model that would fit the combined UK sample data well, according to fit criteria that are used routinely in structural equation modeling. To achieve better model fit, we consulted modification indices (such as LM tests and Wald tests) for adding or removing parameters in the model. We then fitted the modified model and further modified it iteratively until model fit could not be improved further or until the fit was satisfactory. To guard against indiscriminate additions of wastebasket parameters (e.g., error covariances) or factor loadings for the mere purpose of improving model fit,[7] we used some guiding principles to ascertain the reasonableness of the modified CFA model. One principle was that no more than 10% of the total number of possible error covariances be added. Another principle was that the average number of nonzero loadings for a variable not be larger than 2. Translating to the CFA model under consideration, these principles required that no more than 60 error covariances be added for the features and the total number of factor loading parameters be fewer than 70.

After nine iterative model modifications, we obtained a final CFA model. We display the factor pattern in Table 6.7. There are 41 nonzero factor loadings and 35 error covariances in the final model, which increased from 35 nonzero factor loadings and 0 error covariances in the initial model. The final model has a good fit, according to conventional fit criteria: RMSEA = 0.0495, CFI = 0.9134, and SRMR = 0.0828.[8]

In the third stage, we applied the final factor model obtained for the combined UK samples in the second stage to all other countries in a multiple-group analysis setting. Cross-cultural universality would be validated if the

[7] A CFA model can always be fitted perfectly if a sufficient number of error covariances or loadings are added to the model. Adding too many parameters to a CFA model for the sole purpose of improving model fit weakens the scientific value of the hypothesized factor model and is therefore undesirable.

[8] RMSEA is root mean squared error of approximation. An RMSEA value below 0.05 indicates a good model fit. CFI is comparative fit index. A CFA value of 0.9 and above indicates a good model fit. SRMR is standardized root mean squared residual. An SRMR value below 0.05 indicates a good model fit.

Table 6.6 Rotated Factor Pattern with Four Factors

		Factor 1	Factor 2	Factor 3	Factor 4
Memory/memories	(1)	0.68326			
The past	(2)	0.66726			
Remembering	(4)	0.64744			
Personal meaning	(7)	0.60828			
Fond memories	(3)	0.60764			
Reminiscence	(5)	0.57956			
Feeling/emotion	(6)	0.52306			
Thinking	(15)	0.48929			
Childhood/youth	(13)	0.47069			
Happiness	(12)	0.45824		*0.43164*	
Memorabilia/keepsakes	(10)	0.43829			
Rose-tinted memory	(11)	0.33526	*0.33479*		
Wanting to return to past	(18)		0.61046		
Wishing/desire	(20)		0.48149	*0.45238*	
Longing/yearning	(8)	*0.38083*	0.46001		
Reliving/dwelling	(16)	*0.34690*	0.42895		
Comfort/warmth	(19)			0.66205	
Calm/relaxed	(24)			0.63421	
Dreams/daydreaming	(21)			0.57309	
Social relationships	(9)			0.46923	
Prestige/success	(27)			0.42047	
Sadness/depressed	(30)				0.81568
Pain/anxiety	(34)				0.75645
Negative past	(31)				0.71996
Regret	(25)				0.71224
Loneliness	(29)				0.63296
Solitude	(33)				0.62796
Mixed feelings	(22)				0.55341
Lethargy/laziness	(35)				0.51924
Missing/loss	(17)				0.49329
Homesickness	(26)				0.48778
Distortion/illusion	(32)				0.48589
Change	(23)				0.48339
Aging/old people	(28)				0.39556
Sensory triggers	(14)				0.31299

Notes: The numbers in parentheses indicate the prototypicality order of the features in the normed UK sample.

same CFA model fits well to other countries. Specifically, the multiple-group CFA attempted to test the so-called configural invariance hypothesis (Byrne et al., 1989), such that the other countries would have the same factor structural pattern as that specified for the United Kingdom, as depicted in Table 6.7 (and with the same set of error covariances). Given that configural invariance does not require invariance of parameter values across groups, it is a weaker form of invariance.

The first problem encountered in the multiple-group CFA analysis was that Cameroon and Ethiopia did not have positive definite sample covariance matrices. Because non-positive definiteness of the covariance matrices would lead to convergence problems in model estimation, these two countries had to be excluded from the multiple-group CFA. The multiple-group CFA model for the remaining 14 countries did not fit well (RMSEA = 0.1255, CFI = 0.2989, and SRMR = 0.1839). Under a strict hypothesis testing logic, one would have rejected the null hypothesis of cross-cultural universality of nostalgia conceptions. However, in practice, model modifications explore if a more reasonable multiple-group CFA model can be obtained. The aim is to obtain a well fit modified multiple-group model that is not too different from the original CFA model.

In modifying the multiple-group CFA model, we followed similar principles as those applied in the second analytic stage. Unfortunately, all modified models after the second iteration had more than 60 error covariances and all modified models after the third iteration had more than 70 loadings, thus violating some of our predetermined principles. Nonetheless, we continued the model modification process to determine if it was possible to obtain a reasonably well fit model. After the seventh modification attempt, model fit ceased to improve. In the final modified model (i.e., the best one we could achieve), there were 72 factor loadings and 112 error covariances. The fit was poor: RMSEA = 0.1119, CFA = 0.4500, and SRMR = 0.1749.

To show our best effort in adopting the CFA approach, we fit the normed CFA model to individual countries (except for Cameroon and Ethiopia). That is, we tested configural invariance hypotheses (as prescribed by the pattern in Table 6.7) separately for each of the remaining 14 countries. Table 6.8 shows fit statistics for these countries, ordered by the best model fit using the RMSEA fit index. The Netherlands has the best cross-validation fitting, whereas Romania has the worst. The overall impression from these fittings is that the first 5 countries on the list (starting from the Netherlands and up to Greece) offer some supporting evidence for cross-cultural universality of the nostalgia CFA structure. That is, they all have RMSEAs that are smaller than 0.09. Yet, these values are still above the conventional criterion of 0.05. Overall, the CFA

Table 6.7 Factor Pattern of the Normed CFA Model

		Factor 1	Factor 2	Factor 3	Factor 4
Memory/memories	(1)	**			
The past	(2)	**			
Remembering	(4)	**			
Personal meaning	(7)	**			
Fond memories	(3)	**			
Reminiscence	(5)	**			
Feeling/emotion	(6)	**			
Thinking	(15)	**			
Childhood/youth	(13)	**			
Happiness	(12)	**			
Memorabilia/keepsakes	(10)	**			*
Rose-tinted memory	(11)		*		
Wanting to return to past	(18)	*	**		
Wishing/desire	(20)		**	*	
Longing/yearning	(8)		**		
Reliving/dwelling	(16)		**		
Comfort/warmth	(19)			**	*
Calm/relaxed	(24)			**	
Dreams/daydreaming	(21)			**	
Social relationships	(9)			**	
Prestige/success	(27)			**	
Sadness/depressed	(30)		*		**
Pain/anxiety	(34)				**
Negative past	(31)				**
Regret	(25)				**
Loneliness	(29)				**
Solitude	(33)				**
Mixed feelings	(22)		*		**
Lethargy/laziness	(35)				**
Missing/loss	(17)				**
Homesickness	(26)				**
Distortion/illusion	(32)				**
Change	(23)				**
Aging/old people	(28)				**
Sensory triggers	(14)				**

Notes: Retained loadings from the initial CFA model are indicated by double asterisks. The added loadings are indicated by a single asterisk. The numbers in parentheses indicate the prototypicality order of the features in the normed UK sample. CFA, confirmatory factor analysis.

Table 6.8 Model Fit Statistics of the Normed CFA Model for Countries

	RMSEA	CFI	SRMR
Netherlands	0.0732	0.7658	0.1039
USA	0.0806	0.7836	0.1071
Germany[a]	0.0835	0.6738	0.1129
Israel	0.0842	0.7049	0.1211
Greece[a]	0.0893	0.6689	0.1227
Japan[a]	0.0929	0.6356	0.1196
Australia	0.1051	0.6698	0.1428
Poland	0.1075	0.5988	0.1326
India[a]	0.1081	0.6082	0.1247
Turkey[a]	0.1098	0.5950	0.1563
Uganda[a]	0.1116	0.4579	0.1509
China	0.1139	0.5952	0.1269
Chile	0.1145	0.5887	0.1487
Romania[a]	0.1417	0.3613	0.1458

Note: CFA, confirmatory factor analysis; CFI, comparative fit index; RMSEA, root mean squared error of approximation; SRMR, standardized root mean squared residual.
[a]Negative variance estimates or nonpositive definite predicted covariance matrix was present in the solution.

approach did not support cross-cultural universality in the form of configural invariance, let alone stronger universality that requires parameter invariance. Furthermore, as Table 6.8 shows, some model fitting resulted in problematic parameter estimates, although this problem might not be insurmountable if one can obtain more data.

Discussion

We formulated four criteria for establishing cross-cultural universality of complex constructs that are based on prototype theory. To evaluate these criteria, we proposed statistical tests. Also, to illustrate these criteria and associated tests, we presented an illustrative case study that examined the cross-cultural universality of nostalgia conceptions. We then applied exploratory multivariate techniques (cluster analysis and MDS) to classify and understand different cultural patterns so that useful insights could be drawn for future confirmatory studies. Next, we discuss the methodological assumptions of the cultural universality criteria and the relations among these criteria. We then provide a practical guide for applying the criteria and compare our proposed methodology with the traditional approach based on testing measurement invariance in multigroup CFA models, which is coming under increasing scrutiny (Funder, 2020; Gardiner et al., 2019; Ock et al., 2020).

Methodological Assumptions and Prerequisites

The application of the proposed cultural universality criteria must be based on a well-established set of features for the construct of interest in all cultures. That is, the statistical analysis should not have omitted any important central or peripheral features in any cultures. Otherwise, the ordinality of features and the definitions of the central/peripheral feature sets might not be representative in some cultures, rendering statistical results confounded and incommensurate. Therefore, researchers must be able to justify the completeness of feature sets. For example, Hepper et al. (2014) not only instructed participants to rate the 35 nostalgia features but also asked if there were any other features that participants considered important to define nostalgia. If a culture shows that some important features have not been included in the original set, one must pay attention to the peculiarity of that culture: Is it revealing of the genuine uniqueness of this culture, or is it simply due to an omission in the original feature set construction?

This question brings us to a broader point: The techniques we proposed are suited to an *etic* approach to cross-cultural research (i.e., to test the extent to which conceptions of psychological constructs, such as nostalgia, generalize to other cultures; Segall et al., 1998). This approach is standard when examining simultaneously multiple cultures (Hupka et al., 1985; Russell et al., 1989; Schmitt & Allik, 2005). However, complementary investigations using the *emic* approach (i.e., in-depth exploration within each culture from the perspective of its members via different methods) may help identify new features and subtle cultural differences. In the current case, this could present a valuable route to gaining understanding of African conceptions of nostalgia. A related consideration is the diversity of samples. In Hepper et al.'s (2014) investigation, although samples were drawn from countries across five continents with a range of levels of development and industrialization, participants were all university students. The claims of universality can, of course, only be extended to the types of sample included in the study. Future research would do well to vary the education level of participants as well as other characteristics.

Relations Among the Four Cultural Universality Criteria

Assuming that all related central and peripheral features of a construct have been included in the statistical analyses, how should one use the four proposed criteria to evaluate cultural universality? What should one conclude about cultural universality if not all criteria are satisfied? For example, although the ordinality of nostalgia features is strongly supported by the high rank-order correlations in all cultures (Criterion 1), the elevation criterion is only partially satisfied in many, but not all, cultures (Criterion 4). How does one weigh the evidence and interpret the non-consensual results? To answer this question, it is useful to discuss some relations among the proposed criteria so that researchers can make an informed judgment from practical data analysis.

The Ordinality Criterion Is Paramount

The four cultural universality criteria are not of the same theoretical importance and are not entirely independent of each other. In general, the ordinality criterion is critical to prototype theory (except perhaps in the unlikely case that a construct is defined by uniformly prototypical features). Failing the ordinality criterion is fatal: Two cultures cannot have similar conceptions of a given construct if the features are not ordered similarly. Satisfying this criterion is an essential step to establishing cultural universality.

The Elevation Criterion Strengthens the Universality Claim

The elevation criterion of feature sets can be viewed as a stronger version of the ordinality criterion. That is, if all feature sets in two cultures have similar elevations (i.e., prototypicality levels), then the features in the two cultures are expected to be ordered similarly. However, the converse is not necessarily true. Two cultures can have perfectly matching orders in features even when the elevations of the features (or feature sets) are different. In our case study, we observed that many countries satisfy both criteria (e.g., the United States and Greece), and some countries satisfy the ordinality criterion but not the finer elevation criteria (e.g., Romania and Ireland). Whereas the former case would be favorable to infer cultural universality, the latter is inconclusive. Researchers can attribute failure of the elevation criterion to response biases or response sets, if the ordinality criterion is strongly supported. However, such explanations must be further justified.

The Distinctiveness Criterion Provides a Basis to Examine the Elevation Criterion

Whether it is meaningful to check the elevation criterion depends on the distinctiveness of feature sets. If the feature sets are not distinctive in the normed culture, then there is no need to check the elevation criterion in other cultures. There are two main reasons why the feature sets may not be distinctive. First, it could simply be an empirical fact. That is, the construct under investigation could be ambiguous, with features that vary little in prototypicality (but this is supposed to be a rare case). Alternatively, it could be that the features were incorrectly partitioned. Indeed, evidence for cultural universality that is based on the elevation criterion is as strong as the specific partitioning scheme can indicate. Stronger universality claims require more finely partitioned sets.

The central issue, then, is what the correct number of partitioned sets is and how one can construct them. Is the distinction of central and peripheral features good enough to characterize nostalgia conceptions in all cultures, or is a finer level preferable? It is difficult to answer definitively this question, but the statistical methods we proposed can at least suggest exploratory steps. Specifically, one can start with two main feature sets and then examine whether finer partitions are possible. In fact, one need not have sets of the same size. In our case study, we used equal-partitioned sets simply because there were no prior studies to suggest a specific partitioning. Alternatively, one could conduct a cluster analysis on the prototypicality ratings of features, which could yield well-separated partitioned feature sets of different sizes.

Failing the Distinctiveness Criterion Weakens the Claim of Cultural Universality

What if a new culture fails the distinctiveness criterion? If the ordinality criterion is satisfied to some degree, the failure of the distinctiveness criterion means that although the ordering of features in the new culture is similar to that of the normed culture, there might be culture-specific variability in prototypicality rating in the new culture that renders the established feature sets less distinguishable. Hence, failing the distinctiveness criterion weakens the cultural universality by introducing extra culture-specific variability into the prototypicality rating of the affected feature sets.

The Special Role of the Consistency Criterion

Finally, the consistency criterion requires that central features be rated more consistently than peripheral features in all cultures. This criterion is unique in that it pertains to the dispersion, rather than elevation, of feature ratings. However, this criterion can be confounded with the elevation of ratings. For example, in our current case study, we observed ceiling effects or restriction of range; very highly rated features had less variability and, hence, greater consistency. Table 6.3 shows that only Uganda, which has the lowest average rating in central features, has the same consistency (variability) in the central and peripheral features. Table 6.4 shows that most consistency violations in C2/P1 occur in countries with lower average C2 ratings. Therefore, the consistency criterion might echo and supplement other criteria that are related to the elevation of feature ratings. If a central feature set is not rated more consistently than a peripheral feature set, it could imply that the central feature set is actually not rated highly (or representative) enough. Another possibility is that the peripheral feature set might be so irrelevant in a particular culture that people consistently assign very low ratings to those features. Thus, violations of the consistency criterion need to be explored and interpreted carefully.

Steps for Examining the Cultural Universality Criteria in Practice

A Practical Guide to Implementing the Proposed Methods

Given the relations among the proposed criteria for evaluating cultural universality, we summarize a practical guide for setting them:

1. In the normed culture, establish a suitable set of features for the construct in question, and obtain a centrality index (or indices) for each feature (e.g., scale rating, classification speed, and recall frequency). Study the order of features in terms of their centrality to the construct in question, check the trend of standard deviations of the ordered features, and establish an appropriate number of distinguishable feature sets. Different levels of partitioned feature sets can be investigated to test if strong cultural universality can be established.
2. In the new culture, check the ordinality criterion (Criterion 1). If the data fail this criterion, then there is no need for further analysis. Cultural universality cannot be established. Proceed to the next step if the ordinality criterion is satisfied.
3. Check the consistency criterion (Criterion 2). However, if the normed culture does not rate the central features more consistently or the pattern is unclear, then one has to ascertain that those central features are indeed representative enough—this is because highly central features

will ordinarily be rated more consistently (due to a ceiling effect). Once the consistency criterion is established in the normed culture, it is interesting to determine if the new culture shows the same pattern. If the new culture fails the consistency criterion, then researchers might explore why this happened. Is it due to the variability introduced by some outlying cases in rating those central features in the new culture? Is it due to the inclusion of peripheral features that are being judged categorically as irrelevant in the new culture? Proceed to the next step.

4. Check the distinctiveness criterion (Criterion 3). This assumes that the normed culture has already established distinctive feature sets. If the new culture fails to establish the same groups of distinctive feature sets, then the ordinality/elevations of the features might have been confounded by extra culture-specific variability of prototypicality rating in the new culture. If the new culture satisfies the distinctiveness criterion, then the ordinality/elevations of the features have unconfounded interpretations. Proceed to the next step.
5. Check the elevation criterion (Criterion 4). Examine the elevation of each feature set to determine if it matches that of the corresponding elevation of the normed culture. A strong universality claim is established when all elevations match. Stronger universality claims can be established with an increased number of partitioned feature sets. If at least some elevations do not match, report the discrepancies and explore the reasons why. Is it due to response biases or substantive cultural reasons?
6. Cluster analysis can be used to explore possible clusters of cultures that share the same elevation patterns. Multidimensional scaling can enhance understanding of cultural patterns and trends.

The final exploratory step (6) requires clarification. First, we have proposed cluster analysis as an exploratory statistical technique for finding different cultural patterns. The analysis provides a way to group cultures at different levels of clustering, but it does not usually provide a statistical test that enables researchers to determine the correct number of clusters. Hence, cluster analysis results should be interpreted with the aid of MDS results and by checking the cultural universality criteria. Needless to say, substantive theories about cultural patterns are invaluable. Second, our proposed application of MDS is novel in that we do not resort to the use of hypothesized "latent" factors to explain the dimensions. Rather, our strategy was to identify the dimensions by associating them with the observed mean patterns for features. The advantage of using this strategy is that the MDS dimensions are interpreted in more objective terms. The limitation is that this strategy is not a general methodology suitable for all MDS applications. The strategy was possible in our case study because we have a relatively large set of features (35) and a relatively large number of objects (i.e., 18 cultures) in the MDS analysis. A large set of features enables one to form stable partitioned feature sets that serve as the basis of comparisons among cultures. A large number of objects (cultures) increases one's chances of identifying enough data points to contrast the mean patterns in graphs, such as those depicted in Figures 6.6 and 6.7.

Identifying a Normed Culture

The previous section describes steps to study cross-cultural universality of a complex psychological construct. These steps assume that a normed culture has been designated for comparisons with other cultures. In our case study about nostalgia, the normed culture was the United Kingdom because most prior research has been conducted within this specific population. However, it may not always be clear which culture should be designated as the normed culture. We propose a heuristic method for identifying a normed culture, as follows.

First, the features of a complex construct are ranked in each of the k cultures, and rank correlations are computed among all cultures. Then, the average absolute rank correlation is computed for each culture by averaging its absolute correlations with the remaining $k - 1$ cultures. Given that cultural universality pertains to commonality, it is reasonable to designate the culture that has the *maximum* average absolute rank correlation as the normed culture so that it bears maximal similarities with all other cultures. Once this normed culture has been designated, the steps described in the previous section can be carried out to study the cross-cultural universality of the target complex psychological construct.

Beyond Confirmatory Factor Analysis and Invariance Tests

The impetus for developing our new approach stemmed, in part, from concerns regarding the practical and theoretical limitations of the popular CFA approach to analyzing cross-cultural data. Practical problems arise because large-scale cross-cultural studies typically examine many countries (e.g., more than five). As a result, the number of potentially noninvariant parameters would be large, the interpretations would be complicated, and the model fitting process would be cumbersome (Byrne & Van de Vijver, 2010). We illustrated some of these difficulties when we applied the CFA approach in the case study of nostalgia. Indeed, the CFA approach already stalled in the model fitting stage, before more meaningful research questions could be addressed. More generally, its main limitations are that (a) CFA places too much emphasis on model fitting so that the final fitted model is highly susceptible to capitalization on chance (MacCallum et al., 1992) and would include many wastebasket parameters that are difficult to interpret, (b) the parameters in a CFA model do not correspond closely to the prototypicality of features and therefore are not immediately interpretable even in a well-fitted CFA model, (c) multiple-group CFA model fitting is prone to optimization problems and ad hoc adjustments, and (d) multiple-group CFA is computational intensive and time-consuming because of the large search space in model modifications. Although the last limitation itself is not directly related to the analytic quality of CFA, our personal experience has been that practical researchers often would compensate this limitation by adding a lot of wastebasket parameters indiscriminately in the fitting process, thus exacerbating the problems described in the first limitation.

In contrast, the strength of the alternative methods we proposed resides primarily in their suitability to the theoretical foundations of prototype theory. By comparison, the CFA approach encounters two challenges. First, the prototypical features of complex constructs have ordinal structures that the CFA models do not or cannot address. There is no direct implication from prototype theory that factor loadings indicate the prototypicality of features (or items). The situation becomes even more complicated when there is more than one factor for a given construct. Which loadings (or combination thereof) can indicate the prototypicality of features? In fact, Hepper et al. (2014) factor-analyzed the nostalgia features and found that the magnitudes of the loadings did not indicate consistently the prototypicality of features. Second, the most important structural information in prototype theory is that of the mean structures, not the covariance structures. Comparing features among cultures is primarily based on their elevations (i.e., means) that reflect prototypicality. The traditional CFA approach does not consider the mean structures and therefore omits the elevation information altogether. With the advent of multiple-group CFA analysis (Jöreskog, 1971), group differences in the mean structures became more relevant in CFA for cultural data (Byrne et al., 1989). However, the mean parameters in CFA models (i.e., the measurement intercepts and factor means) are still unrepresentative of prototypicality themselves. In contrast, the prototypicality of features can simply be reflected directly by their mean ratings, which are the quantities used in our proposed methodology.

In summary, under prototype theory of complex constructs, it is practically and theoretically problematic to examine cultural universality by using the CFA approach. The methods we propose focus on establishing cultural universality based on the ordinality and elevations of the prototypicality structures. The criteria proposed afford straightforward statistical tests without ad hoc model fitting. Cultures that have non-conformity in features or feature sets are detected readily with standard statistical tests and graphical techniques.

We emphasize that we do not dismiss CFA as a useful methodology in many cross-cultural research situations. Rather, if prototype theory provides a suitable perspective on the complex constructs in question, our proposed methodology is comprehensive and informative. The issue hinges on a critical theoretical distinction between the prototype and CFA approaches. The prototype approach emphasizes the cognitive representations of complex constructs in the form of feature prototypicality, whereas the CFA approach emphasizes the factorial structures of complex constructs in the form of a confirmatory factor model. The consequence is that the prototype approach would claim cultural universality by observing similar cognitive representations in cultures, whereas the CFA approach would claim cultural universality by observing similar functional structures (i.e., factor structures) in cultures. Which approach should be used and under what situations? Can these two approaches be somehow combined and resolved? These are pressing and generative questions for future research. We hope that our proposed method also proves generative by unlocking the potential of a prototype approach to the study of cross-cultural similarities and differences.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. SAGE.

Arabie, P., Carroll, D., & deSarbo, W. S. (1987). *Three-way scaling: A guide to multidimensional scaling and clustering*. SAGE.

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Multivariate Software.

Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology*, *41*, 656–670. https://doi.org/10.1037/0022-3514.41.4.656

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement equivalence. Psychological Bulletin, *105*, 456–466. https://doi.org/10.1037/0033-2909.105.3.456

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*, 107–132. https://doi.org/10.1080/15305051003637306

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*, 155–175. https://doi.org/10.1177/0022022102250225

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. *Advances in Experimental Social Psychology*, *12*, 4–53. https://doi.org/ 10.1016/S0065-2601(08)60258-0

Cantor, N., Mischel, W., & Schwartz, J. (1982). A prototype analysis of psychological situations. *Cognitive Psychology*, *14*, 45–77. https://doi.org/10.1016/0010-0285(82)90004-4

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, *54*, 541–557. https://doi.org/10.1037/0022-3514.54.4.541

Cohen, J. C. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Donald M. (1991). Origins of the modern mind: Three stages in the evolution of culture and cognition. Harvard University Press, Cambridge.

Elshout, M., Nelissen, R. M. A., & van Beest, I. (2017). Conceptualising humiliation. *Cognition and Emotion*, *31*, 1581–1594. https://doi.org/10.1080/02699931.2016.1249462

Fehr, B. (1988). Prototype analysis of the concepts of love and commitment. *Journal of Personality and Social Psychology*, *55*, 557–579. https://doi.org/10.1037/0022-3514.55.4.557

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*, 464–486. https://doi.org/10.1037/0096-3445.113.3.464

Fischer, A. H., Manstead, A. S. R., & Rodriguez Mosquera, P. M. (1999). The role of honour-related vs. individualistic values in conceptualising pride, shame, and anger: Spanish and Dutch cultural prototypes. *Cognition and Emotion*, *13*, 149–179. https://doi.org/10.1080/026999399379311

Fitness, J., & Fletcher, G. J. O. (1993). Love, hate, anger, and jealousy in close relationships: A prototype and cognitive appraisal analysis. *Journal of Personality and Social Psychology*, 65, 942–958. https://doi.org/10.1037/0022-3514.65.5.942

Frei, J. R., & Shaver, P. R. (2002). Respect in close relationships: Prototype definition, self-report assessment, and initial correlates. *Personal Relationships*, *9*, 121–139. https://doi.org/10.1111/1475-6811.00008

Funder, D. (2020, January 31). MIsgivings: Some thoughts about "measurement invariance." *funderstorms*. https://funderstorms.wordpress.com/2020/01/31/misgivings-some-thoughts-about-measurement-invariance

Gardiner, G., Sauerberger, K., Members of the International Situations Project, & Funder, D. (2019). Towards meaningful comparisons of personality in large-scale cross-cultural studies. In A. Realo (Ed.), *In praise of an inquisitive mind: A Festschrift in honor of Jüri Allik on the occasional of his 70th birthday* (pp. 123–139). University of Tartu Press.

Gregg, A. P., Hart, C. M., Sedikides, C., & Kumashiro, M. (2008). Everyday conceptions of modesty: A prototype analysis. *Personality and Social Psychology Bulletin*, *34*, 978–992. https://doi.org/10.1177/0146167208316734

Hauser, M., Chomsky, N., & Fitch, W. (2002). The faculty of language: What is it, who has it and how did it evolve? Science, 298, 1569-1579. doi:10.1126/science.298.5598.1569

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

Hepper, E. G., Ritchie, T. D., Sedikides, C., & Wildschut, T. (2012). Odyssey's end: Lay conceptions of nostalgia reflect its original Homeric meaning. *Emotion*, *12*, 102–119. https://doi.org/10.1037/a0025167

Hepper, E. G., Wildschut, T., Sedikides, C., Ritchie, T. D., Yung, Y.-F., Hansen, N., Abakoumkin, G., Arikan, G., Cisek, S. Z., Demassosso, D. B., Gebauer, J. E., Gerber, J. P., Gonzalez, R., Kusumi, T., Misra, G., Rusu, M., Ryan, O., Stephan, E., Vingerhoets, A. J. J., & Zhou, X. (2014). Pancultural nostalgia: Prototypical conceptions across cultures. *Emotion*, *14*, 733–747. https://doi.org/10.1037/a0036790

Horowitz, L. M, French, R., & Anderson, C. A. (1982). The prototype of a lonely person. In L. Peplau & D. Perlman (Eds.), *Loneliness: A sourcebook of current theory, research, and therapy* (pp. 183–205). Wiley-Interscience.

Hupka, R. B., Buunk, B., Falus, G., Fulgosi, A., Ortega, E., Swain, R., & Tarabrina, N. Y. (1985). Romantic jealousy and romantic envy: A seven-nation study. *Journal of Cross-Cultural Psychology*, *16*, 423–446. https://doi.org/10.1177/0022002185016004002

Hurtado de Mendoza, A., Fernández-Dols, J. M., Parrott, W. G., & Carrera, P. (2010). Emotion terms, category structure, and the problem of translation: The case of shame and vergüenza. *Cognition and Emotion*, *24*, 661–680. https://doi.org/10.1080/02699930902958255

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426. https://doi.org/10.1007/BF02291366

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 user's reference guide*. Scientific Software.

Kearns, J. N., & Fincham, F. D. (2004). A prototype analysis of forgiveness. *Personality and Social Psychology Bulletin*, *30*, 838–855. https://doi.org/10.1177/0146167204264237

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. SAGE.

Lambert, N. M., Graham, S. M., & Fincham, F. D. (2009). A prototype analysis of gratitude: Varieties of gratitude experiences. *Personality and Social Psychology Bulletin*, *35*, 1193–1207. https://doi.org/10.1177/0146167209338071

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504. https://doi.org/10.1037/0033-2909.111.3.490

Matsumoto, D., & Van de Vijver, F. J. R. (2011). *Cross-cultural research methods in psychology*. Cambridge University Press.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–113. https://doi.org/10.1146/annurev.ps.32.020181.000513

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, *112*, 179–204. https://doi.org/10.1037/0033-2909.112.2.179

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.

Muthén, L. K., & Muthén, B. O. (2012). *MPlus user's guide* (7th ed.). Muthén & Muthén.

Ock, J., McAbee, S. T., Mulfinger, E., & Oswald, F. L. (2020). The practical effects of measurement invariance: Gender invariance in two Big Five personality measures. *Assessment*, *27*, 657–674. https://doi.org/10.1177/1073191119885018

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008) Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. Behavioral and Brain Sciences 31(2), 109–178.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Erlbaum.

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, *110*, 426–450. https://doi.org/10.1037/0033-2909.110.3.426

Russell, J. A., Lewicka, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology*, *57*, 848–856. https://doi.org/10.1037/0022-3514.57.5.848

SAS Institute. (2014). *SAS/STAT 13.2 user's guide*.

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, *89*, 623–642. https://doi.org/10.1037/0022-3514.89.4.623

Segall, M. H., Lonner, W. J., & Berry, J. W. (1998). Cross-cultural psychology as a scholarly discipline: On the flowering of culture in behavioral research. *American Psychologist*, *53*, 1101–1110. https://doi.org/10.1037/0003-066X.53.10.1101

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, *46*, 356–360. https://doi.org/10.1016/j.jesp.2009.10.016

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52, 1061–1086. https://doi.org/10.1037/0022-3514.52.6.1061

Shi, Y., Gregg, A. P., Sedikides, C., & Cai, H. (2020). Lay conceptions of modesty in China: A prototype approach. *Journal of Cross-Cultural Psychology*, *52*(2), 155–177. https://doi.org/10.1177/0022022120985318

Uskul, A., Cross, S., Alozkan, C., Gercek-Swing, B., Ataca, B., Gunsoy, C., & Sunbay, Z. (2014). Emotional responses to honor situations in Turkey and the northern USA. *Cognition and Emotion*, *28*, 1057–1075. https://doi.org/10.1080/02699931.2013.870133

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* SAGE.

Van Hemert, D. A., Poortinga, Y. H., & Van de Vijver, F. J. R. (2007). Emotion and culture: A meta-analysis. *Cognition and Emotion*, *21*, 913–943. https://doi.org/10.1080/02699930701339293

Westen, D., Shedler, J., Bradley, B., & DeFife, J. A. (2012). An empirically derived taxonomy for personality diagnosis: Bridging science and practice in conceptualizing personality. *American Journal of Psychiatry*, *169*, 273–284. https://doi.org/10.1176/appi.ajp.2011.11020274