Reference

1. R. J. Charlson, F. P. J. Valero, J. H. Seinfeld, *Science* **308**, 806 (2005).

How to Measure National Stereotypes?

BECAUSE IT IS PARTICULARLY DIFFICULT TO EVALuate the accuracy of national stereotypes, the Report by A. Terracciano *et al.* ("National character does not reflect mean personality trait levels in 49 cultures," 7 Oct. 2005, p. 96) examining the relations between ratings of national character and ratings of individuals in 49 different cultures represents quite a technical achievement. Studies of stereotypes usually suggest that stereotypic beliefs contain a kernel of truth: The perceived differences between groups do in fact exist, but they are smaller than the stereotype would suggest (*1*, *2*). Terracciano *et al.* instead found that, on average, there was no relation between national stereotypes and self and other descriptions. Some methodological weaknesses of their study must be considered, however.

One issue is their almost exclusive reliance on college student samples. Although there is some evidence that cross-cultural comparisons between college students may generalize to broader populations (*3*), there is also substantial evidence that findings with college students frequently do not so generalize (*4*). These findings do not invalidate college student samples as representations of broader national populations, but neither do they justify assuming college students provide an acceptable proxy for the population as a whole.

A second issue is whether the authors have provided a sufficient evaluation of national character. The authors reduce national character to personality traits. This ignores other potential elements of stereotype, most particularly differences in values, beliefs, or perceptions that are not adequately included in the measures used in this study.

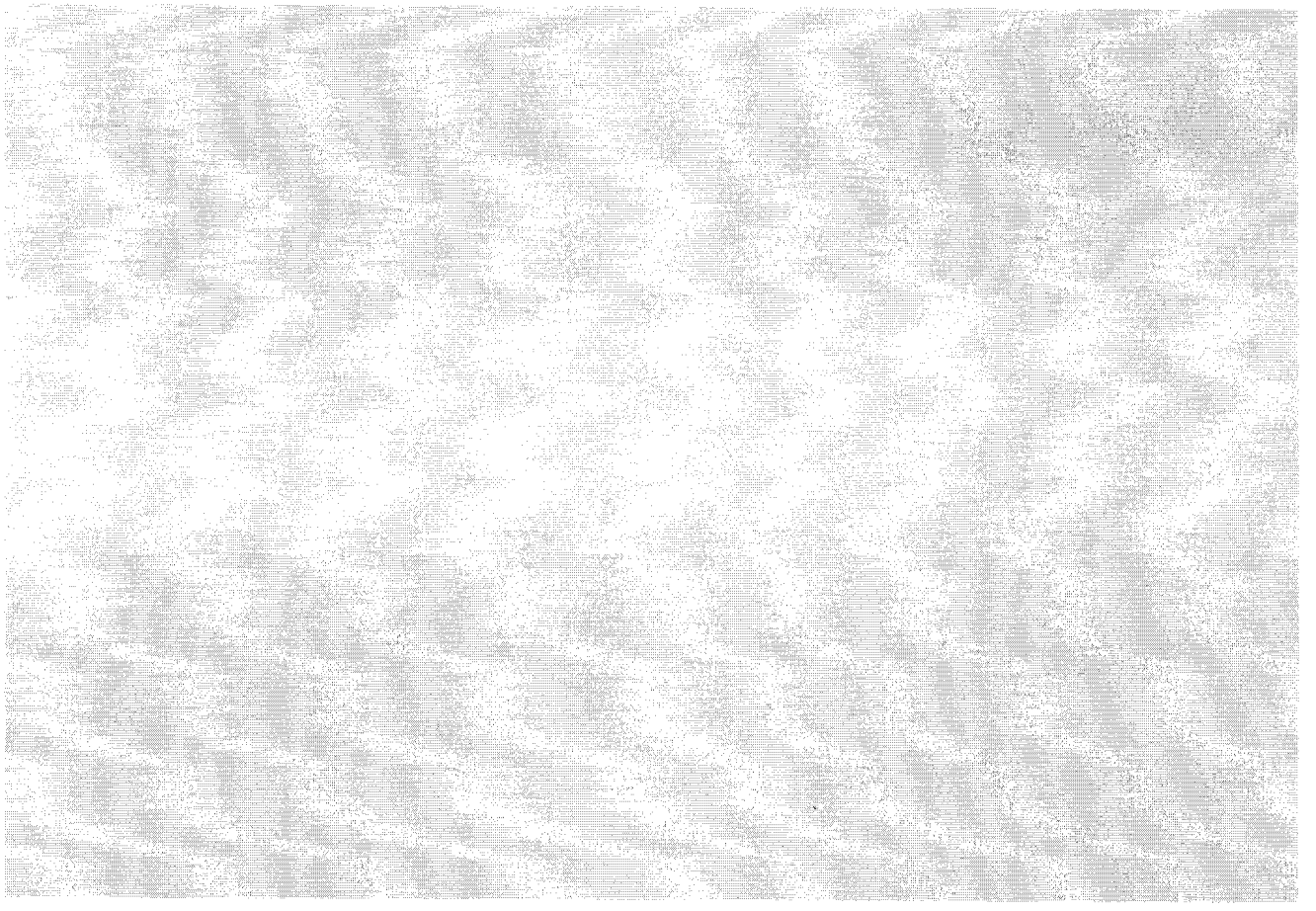
Finally, Terracciano *et al.*'s measures of perceived national character were the mean ratings of the culture by members of that culture. Stereotypes are usually defined in terms of perceptions of the target group by outside observers. Moreover, their measure of actual national character was the mean ratings of oneself or a significant other. In other words, the measurement of national character was based on the ratings of a culture, whereas the measurement of actual character was based on the ratings of a person. The contexts of the two kinds of assessments were quite different and potentially not comparable.

It is increasingly evident that context is an important contributor to outcomes on rating scales (*5*). There is even evidence that cultural differences by themselves can produce differences in the context of the measurement (*6*). A person familiar to the respondent will likely be evaluated in relation to other individuals familiar to the respondent, while a person asked to rate the culture will rate it in relation to other cultures. It is not surprising then to find that these ratings were on average unrelated to ratings of the country's national character.

It is possible that there really is no relation between national stereotypes and actual behaviors. One must wonder, however, what is the source of the variability in the ratings of cultures. Why, for example, do the German Swiss believe they are so conscientious? Even more curious is why Indonesians and Chileans accept that they are not. It seems likely that when asked to rate themselves on conscien-

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.



Published by AAAS

tiousness, German Swiss evaluate themselves in light of those around them. A more definitive test would be to have the German Swiss rated by members of other cultures, but then that is presumably the kernel from which cultural stereotypes germinate in the first place.

ROBERT E. MCGRATH[1]* AND LEWIS R. GOLDBERG[2]

[1]School of Psychology, Fairleigh Dickinson University, 1000 River Road, Teaneck, NJ 07666, USA. [2]Oregon Research Institute, 1715 Franklin Boulevard, Eugene, OR 97403–1983, USA.

*To whom correspondence should be addressed. E-mail: mcgrath@fdu.edu

References

1. J. S. Hyde, Am. Psychol. 60, 581 (2005).
2. H. M. Coon, M. Kemmelmeier, *J. Cross-Cultural Psychol.* **32**, 348 (2001).
3. S. H. Schwartz, A. Bardi, *J. Cross-Cultural Psychol.* **32**, 268 (2001).
4. R. A. Peterson, *J. Consumer Res.* **28**, 450 (2001).
5. N. Schwarz, *Am. Psychol.* **54**, 93 (1999).
6. N. Schwarz, *J. Consumer Res.* **29**, 588 (2003).

CONSISTENT WITH A LONG-HELD VIEW IN SOCIAL psychology, A. Terracciano and colleagues claim that national stereotypes lack accuracy ("National character does not reflect mean personality trait levels in 49 cultures," Reports, 7 Oct. 2005, p. 96). Although it is possible that their findings demonstrate people's inability to discern the attributes of their own groups, three alternative explanations need to be considered.

First, the criterion scores, which were obtained from responses on a personality inventory [the Revised NEO Personality Inventory (NEO-PI-R)], were less variable than the stereotype scores, which were obtained with a new instrument [the National Character Survey (NCS)]. Arguably, the greater length of the NEO-PI-R facet scales (eight items) relative to the NCS scales (one item) contributed to this difference. Furthermore, the nonrepresentative sampling of respondents could have reduced the variability of the criterion scores, as college students tend to share similarities in different cultures.

Second, the similarity of the sample profiles was assessed with intraclass correlation coefficients (ICCs). ICCs are used for dyadic data that cannot be sorted. When judgments are correlated with criteria, Pearson correlations are more appropriate. These indices are only sensitive to profile similarity, not to differences in variability.

Third, national characteristics and stereotypes can be specific. The Japanese may be uniquely characterized by their deference, whereas people from the United States may be known for their materialism. If so, measures of profile similarity gravitate toward zero as a function of profile length.

Failures to reject a null hypothesis are usually not newsworthy. A typical response is to design a study to minimize contaminating effects. Here, however, the embrace of the null hypothesis is also a conceptual surprise. Historically, research on the five-factor model of personality has been predicated on observer agreement, where agreement was thought to imply accuracy. Now, the role of observer agreement is to signal inaccuracy. It is certainly possible that perceptions of nations are qualitatively different from perceptions of individuals, but to find out we need a process model that specifies how people judge national character and how they might agree without being accurate.

JOACHIM I. KRUEGER AND JACK C. WRIGHT

Department of Psychology, Brown University, Hunter Laboratory 295, 89 Waterman Street, Providence, RI 02912, USA.

Response

WE AGREE WITH MCGRATH AND GOLDBERG THAT national stereotypes include more than national character, and beliefs about national differences in appearance, attitudes, or athletic abilities may or may not be accurate. Our study focused on personality traits, which seem to define the core of national character. To the extent that the five-factor model (FFM) is comprehensive, our National Character Survey (NCS) measured key features of national character, and we found no evidence for a kernel of truth in these stereotypes.

Student samples may or may not generalize to