

What's Wrong With Cross-Cultural Comparisons of Subjective Likert Scales?: The Reference-Group Effect

Steven J. Heine and Darrin R. Lehman
University of British Columbia

Kaiping Peng
University of California, Berkeley

Joe Greenholtz
University of British Columbia

Social comparison theory maintains that people think about themselves compared with similar others. Those in one culture, then, compare themselves with different others and standards than do those in another culture, thus potentially confounding cross-cultural comparisons. A pilot study and Study 1 demonstrated the problematic nature of this reference-group effect: Whereas cultural experts agreed that East Asians are more collectivistic than North Americans, cross-cultural comparisons of trait and attitude measures failed to reveal such a pattern. Study 2 found that manipulating reference groups enhanced the expected cultural differences, and Study 3 revealed that people from different cultural backgrounds within the same country exhibited larger differences than did people from different countries. Cross-cultural comparisons using subjective Likert scales are compromised because of different reference groups. Possible solutions are discussed.

Much research in cultural and cross-cultural psychology as well as in social and personality psychology relies on comparisons of means across groups of self-report measures of attitudes, traits, and values. We question the validity of such comparisons.

Although there are a variety of approaches for studying psychology and culture (for reviews, see Greenfield, 1997; Nisbett & Cohen, 1996; Shweder et al., 1998; Triandis, McCusker, & Hui, 1990), one of the most widely used strategies has been to contrast cultures on the basis of *cultural syndromes*, which are patterns of shared attitudes, beliefs, or values that are organized around a theme and largely shared by members of an identifiable group (Triandis, 1996). Examples of such syndromes are tightness and complexity (Triandis, 1989), mastery and conservatism (Schwartz, 1994), and power distance and uncertainty avoidance (Hofstede, 1980). Research guided by this approach has sought to compare or rank order cultures on the basis of their positions on these dimen-

sions. The value of such an approach is compelling: If we can identify the variables by which cultures differ, we can conceivably map out the cultures of the world and create a universal psychology that incorporates each culture's indigenous psychology (Triandis, 1996, p. 407).

Individualism–Collectivism

Among the growing list of cultural syndromes under investigation, one construct stands out above the rest in terms of stimulating research: individualism–collectivism (IC). IC has been conceptualized as a single dimension (e.g., Hofstede, 1980; Kiuchi, 1995), as two dimensions (e.g., Kagitcibasi, 1994; Singelis, 1994), as three dimensions (e.g., Rhee, Uleman, & Lee, 1996), and as four dimensions (e.g., Takata, 1999; Triandis, 1996). This construct is sometimes referred to as independent versus interdependent views of the self (Markus & Kitayama, 1991b; Singelis, 1994), idiocentrism–allocentrism (Triandis, 1989), or agency–communion (Bakan, 1966). Each of these definitions subtly differs from the others, but they converge in their description of an orientation toward focusing on oneself either as an individual or as a member of a significant in-group. In this article we refer to this construct as IC. Triandis (1989) defined individualists as those who “give priority to personal goals over the goals of collectives” and collectivists as those who “either make no distinctions between personal and collective goals, or if they do make such distinctions, they subordinate their personal goals to the collective goals” (p. 509). The present article is focused on methodological concerns regarding cultural comparisons of IC, although the concerns raised are issues for many comparisons of attitude, trait, and value scales across cultures.

It is important to note that much psychological theory has discussed how different cultures vary in the degree to which they

Steven J. Heine, Darrin R. Lehman, and Joe Greenholtz, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada; Kaiping Peng, Department of Psychology, University of California, Berkeley.

Preparation of this article was supported by a fellowship from the Japan Foundation, National Institute of Mental Health Grant R01 MH060155-01A2, and start-up funds from the University of Pennsylvania and the University of British Columbia to Steven J. Heine; and by grants from the Social Sciences and Humanities Research Council to Darrin R. Lehman. We are especially grateful to Yu Niiya and Nariko Takayanagi for their help in the translations.

Correspondence concerning this article should be addressed to Steven J. Heine, Department of Psychology, 2136 West Mall, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. E-mail: heine@cortex.psych.ubc.ca

adhere to, elaborate, afford, and construct experiences associated with IC. Specifically, Western, and especially North American, cultures are typically characterized as scoring high in individualism, whereas cultures in the rest of the world, particularly East Asia, are usually described as embracing the values of collectivism (e.g., Markus & Kitayama, 1991b; Triandis, 1989). These hypothesized cultural differences in IC have been recruited to explain cultural differences in various psychological processes between East Asians and North Americans (for reviews, see Heine, 2001; Markus & Kitayama, 1991b).

Assessing Cultures

Much research under the aegis of IC theory has involved attempts by researchers to measure cultural attitudes and values related to IC. The approach typically taken is a variant on how psychologists measure personality. Participants are asked how much they believe or endorse items on subjective Likert scales, such as traits, beliefs, attitudes, values, or behavioral statements that are believed to reflect patterns of IC. These scores are then compared across cultures or are first subjected to some kind of data reduction technique, such as factor analysis, before the cultural groups are compared on their factor scores.

Cross-cultural comparisons are rife with methodological confounds that the literature has addressed in detail. For example, there are potential cultural differences in moderacy response styles, in which people from one culture are more likely to answer toward the center of a scale than are people from another (Chen, Lee, & Stevenson, 1995). Translation errors can produce measures conveying different meanings across cultures (Brislin, 1970). Members of some cultures may have more familiarity with completing questionnaires than do others (Greenfield, 1997). It is plausible that members of some cultures may be more likely to disguise their responses behind a façade of modesty than are others (Heine, Takata, & Lehman, 2000). These are all significant concerns in the field and have been discussed at length elsewhere. For the most part, it appears that various efforts, such as back-translation (Brislin, 1970), the use of different methodologies on the basis of the nature of the sample (Greenfield, 1997), the use of hidden behavioral measures (Heine et al., 2000, 2001), and standardization of the data (Bond, 1988), can correct many of these confounds.

Peng, Nisbett, and Wong (1997) raised another question about such cross-cultural comparisons: What is the criterion of validity? Indeed, many of these studies are conducted without any basis of validity other than the face validity of the items. Without a solid criterion, these studies risk yielding invalid and misleading results. Peng et al. provided one such potential validity criterion: cultural experts of the cultures being compared. They found that comparisons of values between Chinese and Americans did not match the cultural experts' predictions. The problem was not that the value scales accounted for only a small portion of the variance; rather, the value measures accounted for virtually none of the variance in the cultural experts' predictions.

Peng et al. (1997) offered two explanations for the divergence between participants' ratings and the experts' assessments, although they did not offer an empirical test of either explanation. First, they suggested that a deprivation model might be operating. That is, people come to value what they do not have. Just as those

who are hungry are especially likely to value food (Maslow, 1943), people from cultures that feel deprived of certain needs come to value them more. Peng et al. noted that Singaporean Chinese are more likely than Americans to value choosing their own goals, and Americans are more likely than Chinese to value humility. Simply inspecting these data leads one to conclude that Singapore is more a culture of freedom of choice and America is more a culture of humility, even though such a conclusion challenges widely shared stereotypes of the two cultures. The deprivation effect seems especially problematic for comparisons of values and is less an issue for comparisons of traits (i.e., people who feel particularly lacking in a trait are unlikely to evaluate themselves highly on that trait). Although we do not explore the role of deprivation in the present studies, we note that it, too, is a pernicious problem in the assessment of culture.

Second, Peng et al. (1997) noted that people from different cultural groups use different referents in their self-reported values. For example, Chinese evaluate themselves in comparison with other Chinese, whereas Americans evaluate themselves with reference to other Americans. We term this the *reference-group effect*, and the present article examines this confound in cross-cultural comparisons.

The Reference-Group Effect

The reference-group effect is the confounding role of context in comparisons of mean questionnaire responses across different groups, in particular (but not exclusively) across different cultures. Although the notion that one's reference group affects perceptions has not been considered in studies comparing cultures, it has been an important topic in social psychology for decades. For example, Sherif (1936) demonstrated that people come to view the apparent movement of a light in a room largely on the basis of how their reference group is viewing it. Hyman (1942) highlighted how one's status on a particular dimension is determined by an understanding of oneself within the context of one's reference group. Latane and Darley (1970) demonstrated that how a potential emergency is perceived depends heavily on the reactions of those around one.

The role of others in people's perceptions of themselves was most clearly articulated by Festinger (1954), in his landmark theory of social comparison processes. Festinger maintained that much of people's understanding of themselves is not context-free but rather is based on how people compare with others around them. For example, there is no universal consensus on what the objective threshold is for being considered tall—rather, being tall depends on comparisons with appropriate targets, typically those of similar age, sex, and nationality. The same height—for example, 5 feet 9 in.—is seen as tall in some contexts (e.g., among elementary school children or Japanese women) and short in others (e.g., among professional basketball players or Dutch men). Social comparisons tend not to be effortful, consciously controlled processes. Rather, they are made spontaneously and unintentionally and may affect one's self-evaluations without any deliberate awareness (Baldwin & Holmes, 1987; Latane & Darley, 1970). The mere presence of others leads people to evaluate themselves in reference to those others.

When social comparison targets are chosen, they are not chosen at random. Some targets are chosen to allow the individual to

self-enhance (R. L. Collins, 1996; Wills, 1981), and others are chosen to gain information about one's standing (Festinger, 1954; Wood, 1989). It is important to note that social comparison is largely directed towards similar others, as it is more difficult to gain an accurate appraisal by comparisons with dissimilar others (Festinger, 1954). For example, an American college student evaluating herself on extraversion is not likely to base her self-evaluation on how she compares with the Dalai Lama. Rather, she is likely to compare herself with her friends or peers, who might share her dorm or be taking the same classes as she is.

Biernat and colleagues (Biernat & Billings, in press; Biernat & Kobrynowicz, 1997; Biernat & Manis, 1994; Biernat, Manis, & Nelson, 1991) have proposed that the integral role of context in evaluation leads to shifting standards in evaluations of stereotyped groups. Specifically, people make evaluations of a target by contrasting the individual with other members of that target's group, not with the population as a whole. To the extent that two groups differ in their average level on the dimension under question, the groups have different standards by which members of those groups are evaluated, which thus confounds comparisons of them. For example, a professor might refer to a colleague as not very bright. What she means, of course, is that the target is not very bright by the standards expected of professors, at least in her opinion. If, by an unfortunate twist of fate, however, this colleague ended up in prison, it would not be surprising if his fellow prisoners viewed him as a genius. The standard for what is seen as intelligent likely diverges between professors and prisoners.

The shifting standards effect is particularly pronounced when responses are measured with subjective Likert scales (Biernat et al., 1991). Likert scales capture one's feelings relative to a comparison group or shared norm, but they do not provide a context-free assessment of one's absolute standing. Rather, as Volkman's (1951) rubber band model proposes, people set the endpoints of Likert rating scales to match the range of the dimension that they expect. For example, when one is evaluating whether a man is tall, the endpoints most likely capture a higher range than they do when one is evaluating a woman. To the extent that there are real differences on a dimension across groups, the use of subjective Likert scales is likely to conceal them. To be most accurate, then, strongly endorsing an item on a Likert scale related to a trait such as extraversion, an attitude such as whether expressing a contrary opinion is good, or a value such as individual freedom reflects how one feels about extraversion, expressing contrary opinions, and individual freedom relative to norms of one's reference group. One solution for the problem of shifting standards is to use more objective response scales. Comparisons of objective and subjective response scales have demonstrated that the former are more resistant to context and contrast effects and thus better preserve the meaning across situations and groups (e.g., Biernat & Manis, 1994; Biernat et al., 1991; Campbell, Lewis, & Hunt, 1958; Krantz & Campbell, 1961; Manis, 1967; Peng et al., 1997).

The shifting standards effect has been explored for evaluations of others, but it is no less problematic for self-assessments. Americans might consider a woman who is 5 feet 9 in. to be tall but not a man of the same height. Likewise, an American woman who is 5 feet 9 in. is likely to view herself as tall, whereas an American man of the same height likely would not. Self-evaluations are made relative to the standards that individuals believe are applicable to them.

Culture and Reference-Group Effects

The integral role of context in self-evaluations is especially problematic for cross-cultural comparisons because, simply put, cultures differ in their membership. For example, Japanese culture is made up of mostly Japanese individuals, whereas Canadian culture is made up of mostly Canadian individuals. That people evaluate themselves on the basis of implicit comparisons with those around them means that Japanese tend to evaluate themselves on the basis of how they compare with other Japanese, whereas Canadians tend to evaluate themselves on the basis of how they compare with other Canadians. Clearly, Japanese do not habitually think about themselves in comparison with Canadians. People from different cultures have different reference groups.

To the extent that the different reference groups in different cultures have the same norms for a particular dimension, cultural comparisons are not confounded. Hence, people from different cultures are comparing themselves with the same standard. For example, if the question is "How good is your hearing?" and the standard of what entails normal hearing is identical across the world, comparisons of responses from one culture with those from another will not be confounded. However, if the cultures vary considerably in the norms associated with a particular dimension, cross-cultural comparisons will be confounded. For example, imagine that the average height for men in Culture A is 5 feet 8 in., whereas the average height for men in Culture B is 6 feet. A man who is 5 feet 9 in. from Culture A would likely feel a little tall (he is above average in his culture), whereas a man who is 5 feet 11 in. in Culture B would likely feel a little short (he is below average in his culture). The different norms in the two cultures thus distort a comparison of people's responses in the direction precisely *opposite* to that of the real differences; the 5 foot 9 in. man from Culture B would feel taller than the 5 foot 11 in. man from Culture A. Moreover, the more cultures differ in their norms (i.e., the more cultures are really different on the dimension), the more the cultural comparisons are confounded. That an important goal of cross-cultural research in psychology is explicitly to assess the norms in different cultures renders the reference-group effect problematic indeed.

We use height as an example because it is a concrete, easy to measure dimension. The confounding effects of different reference groups for cultural comparisons seem to be at least as problematic for measurement of less concrete dimensions; for example, feelings of independence. Many researchers have argued that North Americans are more likely to possess independent construals of self than are Japanese (e.g., Heine, Lehman, Markus, & Kitayama, 1999; Markus & Kitayama, 1991b; Triandis, 1989). For the purpose of illustration let us assume that these researchers are correct: North Americans really are more independent than Japanese are. However, if we were to measure independence in the two cultures using subjective Likert scales and respondents based their evaluations on the norms for independence in their own culture, then the expected range of independence would begin and end at a higher level for North Americans than it would for Japanese (Biernat et al., 1991). Hence, a Japanese individual who endorses the response option *strongly agree* for a question regarding his or her independence might be referring to a lower threshold of independent behavior than is a North American individual. The items have come to take on different meanings across cultures in the direction

that reduces the magnitude of the measured cultural difference. This suggests that we should see less evidence of cultural differences when we assess them with subjective Likert scales than when we assess them with more objective measures.

Indeed, different methodologies do differ in the extent to which they reveal cross-cultural differences. For example, Americans in the southern and northern states show relatively weak cultural differences in attitude measures regarding their feelings about the appropriateness of physical aggression (Nisbett & Cohen, 1996). However, archival evidence, field studies, physiological measures, and behavioral measures reveal more pronounced differences (Nisbett & Cohen, 1996). Measurements of beliefs in the mutability of the self using subjective Likert measures often fail to reveal a cultural difference between East Asians and North Americans (Chiu, Hong, & Dweck, 1997; Heine et al., 2001). However, measures that avoid subjective Likert options reveal far more beliefs in the malleability of the self among Japanese than among Americans (Heine et al., 2001). Similarly, subjective measurements of values among Chinese and Americans do not reflect the predictions of cultural experts, although measurements of the same values using concrete scenarios do (Peng et al., 1997). In sum, subjective measures less successfully identify cultural differences than do more objective methods.

Reference-Group Effects and Cultural Comparisons of IC

The most well-known comparison of cultures is Hofstede's (1980) investigation of cultural values among IBM employees from 40 different countries. Participants answered a large number of work-related questions, and the data were reduced to reveal four correlated factors: Individualism, Power Distance, Uncertainty Avoidance, and Masculinity. Hofstede's findings with the individualism dimension, in particular, sparked confidence that this methodology made it possible to map out the cultures of the world. His analyses revealed that the United States was the most individualistic country of the 40 surveyed, followed, in order, by various English speaking countries, other European countries, and various Latin American and Asian countries. The data nicely fit with people's intuitions of how the world's cultures differ. Hofstede's findings (again, particularly his individualism data) have received much attention. For example, his book *Culture's Consequences* has been cited over 2,000 times. Indeed, no empirical work is more influential in the field of culture and psychology.

The validity of Hofstede's (1980) individualism scores has been demonstrated through a number of methods. For example, E. S. Kashima and Kashima (1998) found that Hofstede's individualism scores correlated significantly with whether the dominant language in that country allowed for pronoun drops. Likewise, Suh, Diener, Oishi, and Triandis (1998) have found that Hofstede's IC scale correlates with whether subjective well-being in a country hinges more on positive feelings or on fulfillment of norms. Also, Hofstede (1980) found that countries with similar cultural histories tend to cluster together. Although these studies provide evidence that the rough ordering of the countries in terms of their individualism scores is valid, it does not necessarily follow that the method used to derive this order is also valid. Indeed, Hofstede's methodology has not escaped criticism (e.g., Takano & Osaka, 1999). We briefly summarize the methodology used by Hofstede to measure individualism across cultures.

Hofstede's original questionnaire that went to IBM employees contained 183 items regarding the work environment, including 63 items earmarked for cross-cultural comparisons. Fourteen of these 63 items were used for the individualism scale, and the resultant individualism index primarily loaded on the following 6 items, reported verbatim as follows (with loadings in parentheses): "Have a job which leaves you sufficient time for your personal or family life" (.86), "Have training opportunities (to improve your skills or learn new skills)" (-.82), "Have good physical working conditions (good ventilation and lighting, adequate work space, etc.)" (-.69), "Fully use your skills and abilities on the job" (-.63), "Have considerable freedom to adapt your own approach to the job" (.49), and "Have challenging work to do—work from which you can get a personal sense of accomplishment" (.46).

How these items relate to the rich theoretical description of IC provided by Triandis (1989), Markus and Kitayama (1991b), and others is not clear and has also been questioned by Takano and Osaka (1999). How does desiring good physical working conditions or training opportunities make one more collectivistic? Why would being able to fully use one's skills be associated with less individualism? It also is worth noting that a couple of the other items used in the 14-item individualism scale that appear to more closely reflect the theoretical literature on individualism did not load highly on the individualism scale: for example, "Work with people who cooperate well with one another" or "Have a good working relationship with your manager."

Two other methodological concerns regarding the development of Hofstede's (1980) individualism scale deserve comment. First, this index is not orthogonal to the other three dimensions that Hofstede developed. For example, it has a correlation of $-.67$ with the Power Distance index, which refers to participants' emotional dependence on more powerful people (Hofstede described this as distinct from his Individualism index's referral to the emotional dependence on groups of people). This theoretical distinction is not altogether clear to us (especially given that the items from Hofstede's individualism scale do not refer to emotional dependence on groups), and the high correlation between these two scales suggests to us that separating these two factors is not empirically supported. Second, the factor analyses that produced the individualism scale were based on 40 observations, namely, the average scores for each country. Factor analyses are notoriously unreliable with small sample sizes, and this adds to our caution in interpreting the obtained pattern of results. In sum, we are concerned by several methodological shortcomings in Hofstede's development of his individualism scale—shortcomings that are rarely addressed despite the huge influence this index has on the field (for an exception, see Takano & Osaka, 1999). We suspect that Hofstede's individualism scale has become so popular because the obtained pattern of results so nicely matches people's intuitions about the ordering of the cultures (i.e., the United States as the most individualistic, Latin America and East Asia as the most collectivistic) and that people assume that with a more rigorous methodology the results would have looked even better.

Schwartz and colleagues (Schwartz, 1994; Schwartz & Bilsky, 1990; Schwartz & Sagiv, 1995) built on Hofstede's (1980) approach, correcting some of the shortcomings inherent in Hofstede's original methodology. Schwartz and colleagues went beyond values in the workplace and included a comprehensive list of 56 different values believed to be recognized in all cultures.

They used a sample likely to be more representative of the world's population than are IBM employees, namely, school teachers and university students. And they used both individual- and culture-level analyses, thus solving the small sample size problem of Hofstede's exclusively culture-level analyses. Schwartz (1994) described seven correlated dimensions: Conservatism, Hierarchy, Mastery, Affective Autonomy, Intellectual Autonomy, Egalitarian Commitment, and Harmony.

However, the ordering of the various countries on these seven dimensions does not seem to us as intuitively satisfying as the findings obtained from Hofstede's (1980) individualism scale. For example, the Affective Autonomy Scale, which includes the values "Enjoying life," "Pleasure," "Exciting life," and "Varied life," finds East Germany scoring the 3rd highest out of 38 countries and Italy scoring the 2nd lowest. The Conservatism Scale, consisting of such values as "Honoring elders," "Politeness," "Self-discipline," and "Clean," finds the United States ranking ahead of Japan. In addition, the Mastery Scale, consisting of such values as "Choosing own goals," "Independent," and "Daring," finds Chinese samples endorsing these more than any other culture in the world (Schwartz, 1994). These findings differ from some commonly held stereotypes of these countries (although there are, of course, other examples that are consistent with various stereotypes). Whether the stereotypes are accurate or even widely shared is a matter of contention, but without a clear criterion of validity it is impossible to evaluate how well Schwartz's sophisticated methodology yields an accurate picture of the world's cultures. The only criterion offered by this research is that countries with similar cultural backgrounds tend to cluster together. However, this pattern is also consistent with expectations given reference-group effects; similar cultures should have similar reference groups. That is, highly individualistic cultures have highly individualistic reference groups, and their responses might thus be based on a similar standard.

One might think that to the extent that people respond to questions about themselves on the basis of a comparison with a cultural norm, there should never be any consistently observed cultural differences. The thinking would be that any differences that exist are obscured because people's responses across cultures average out when compared with their respective referents. However, this would only be the case if people's responses exclusively indicated their assessments relative to their cultural norm. Social comparison is clearly not the only process by which people come to understand themselves; it is merely one important process. It seems that responses to items also involve some introspection that is relatively unaffected by social comparison, a point we return to later. Moreover, in the case of Hofstede's (1980) results, it might be the case that participants were comparing themselves with a common referent, IBM's managerial style, which would have led to the robust and consistent ordering of cultures that Hofstede found.

One distinct advantage of cross-cultural psychology—its ability to consider many different cultures simultaneously—belies an important disadvantage. With so many cultures explored at once, it is difficult to evaluate how well Schwartz's (1994) and Hofstede's (1980) approaches actually map out the world. No individual is expert on all 40 cultures, and there is no literature outside of this work that attempts to place each culture in the context of all of the other cultures of the world. Hence, no one is in the position of

assessing the validity of the impressive set of data amassed through these surveys. This lack of a validity criterion coupled with theoretical concerns regarding reference-group effects and other confounds raises questions about the validity of such cross-cultural comparisons.

There have been a number of recent papers that have reviewed various past studies of IC among Eastern and Western cultures. For example, Matsumoto (1999) and Takano and Osaka (1999) surveyed a subset of the literature on cross-cultural comparisons of trait and attitude measures of IC conducted with Japanese and Americans and noted that there was no difference between the two cultures. Oyserman, Coon, and Kimmelmeier (2002) conducted a thorough review of the database on cross-cultural comparisons of IC worldwide. Most relevant to the present article, Oyserman et al. found little evidence of cultural differences between East Asians and Americans. It is interesting that Japanese and Koreans were just as collectivistic as Americans, and although Americans scored higher than these two countries on individualism, the effect sizes were small (the cultural differences between various Chinese samples were larger).

Pilot Study: A Review of IC Studies Conducted in East Asia and North America

Prior to encountering the Oyserman et al. (2002) paper, we conducted our own review of IC studies that compared East Asians and North Americans.¹ We searched PsycINFO with the keywords *collectivism, individualism, independent, or interdependent* and the keywords *Asian, Japanese, Chinese, Korean, Hong Kong, or Taiwan*. We omitted dissertations and investigations of Asian Americans (it is not clear how acculturated these samples are). There are likely relevant published articles that did not appear in our search, either because they were not listed in PsycINFO with the keywords that we selected, because it was not evident from the abstracts whether the appropriate comparisons were being made, or because the articles did not report the means of the scales being compared. Moreover, we were unable to locate a few of the articles that were listed in PsycINFO. Our search should thus be viewed as a systematic but imperfect attempt to survey the most accessible published work in this area.

Our review uncovered 76 distinct cross-cultural comparisons of subscales relating to IC (a summary of this is available on request). Given the diversity of opinions regarding how many dimensions exist in the IC construct, we made all comparisons at the subscale level. We noted whether the direction of the subscale means was consistent with what Takano and Osaka (1999) labeled the *common view*. The common view is the belief that East Asians score higher than North Americans on subscales relating to collectivism and that North Americans score higher than East Asians on subscales relating to individualism. We then calculated a sign test of the number of comparisons that were in the direction of the common view versus those that were not. Chance alone predicts that 38 of these comparisons should be in the direction consistent with the common view. In actuality, 42 of the comparisons were consistent with the common view, and 34 were in the opposite

¹ This review was conducted in summer of 1999 and, hence, does not include articles published after that.

direction. A sign test reveals that this is not greater than chance ($z = .80, ns$). That is, similar to the findings of other researchers (e.g., Matsumoto, 1999; Oyserman et al., 2002; Takano & Osaka, 1999), we found no evidence of cultural differences in IC between East Asian and North American samples. This null pattern is not owing to a restricted measurement of IC, as 21 different measures were included in this analysis.

The four meta-analyses converge to show that there are no consistent cultural differences between East Asian and North American samples on trait, attitude, and value measures of IC. There is thus a striking divergence between the theory on cultural differences of IC and the data from attitude, trait, and value measures.

Study 1: IC Items Compared With a Validity Criterion

The findings from the meta-analyses are puzzling. We remind readers that IC is the most discussed dimension in the literature of culture and psychology, and the two poles of this dimension are often assumed to be occupied by North Americans and East Asians. What is the field to make of the null pattern from these meta-analyses? Is this large theoretical literature on IC grossly mistaken?

Before saying farewell to the common view, we consider evidence from other disciplines concerning the notion of greater collectivism in East Asia relative to North America. Strong collectivistic tendencies among East Asians have been identified by anthropologists (e.g., Hendry, 1987; Kondo, 1990; Lebra, 1976; Rosenberger, 1992), education researchers (e.g., Ichiki, 1985; Lewis, 1995; Tobin, Wu, & Davidson, 1989), Japanologists (e.g., Hamaguchi, 1985; Reischauer, 1988), journalists (e.g., Buruma, 1984; Christopher, 1983; Whiting, 1990), philosophers (e.g., Nakamura, 1964), psychiatrists (e.g., Doi, 1971; Johnson, 1993; Roland, 1988), sociologists (e.g., Nakane, 1970; Vogel, 1979), and the popular press (e.g., R. J. Collins, 1992; Feiler, 1991; Miyamoto, 1994). The terms *collectivistic* or *interdependent* are not always used (indeed, many researchers are averse to such sweeping labels and generalizations of cultural differences); however, these other sources of evidence are consistent in describing a greater concern with relationships, interpersonal harmony, role-appropriate behavior, and fulfillment of obligations to others for East Asians compared with North Americans (but note that some researchers have taken issue with the common view; e.g., Befu, 1980; Mouer & Sugimoto, 1986).

However, selective reports and anecdotes gleaned from such diverse literatures and with such variegated and abstract concepts as individualism and independence may be too subjective and potentially biased a methodology for psychologists' tastes to count as evidence. In Study 1, we sought to secure a more empirical basis for this comparison by asking Japanese specialists living in North America to indicate whether they believe that various thoughts and behaviors are more or less common among Japanese than they are among North Americans. We examined whether there is a consensus among cultural experts regarding cultural differences in the kinds of specific thoughts and behaviors that make up the scales that did not discriminate between cultures in past studies of IC.

Method

Participants. We asked all of the faculty who specialize in topics related to Japan at the University of Pennsylvania, the University of California, Berkeley, and the University of British Columbia to complete a brief questionnaire. A total of 14 agreed to do so (none of whom are psychologists).

Materials. There are many different measures of IC that we could have used. However, as we aimed to keep the questionnaire brief to encourage participation from our cultural experts, we decided to use just one IC measure. We selected Singelis's (1994) Independence and Interdependence Scale for the study largely because it has sound psychometric properties, has high theoretical overlap between the items and the literature on IC, and is one of the most used scales for measuring the construct. Our cultural experts were presented with the scale, and they were asked to consider the various Japanese and North Americans whom they knew personally, through their research, through casual observations, from television, and so forth.² For each of the 23 items (12 interdependent items and 11 independent items), they were asked to indicate whether the item was more characteristic of Japanese than of North Americans or more characteristic of North Americans than of Japanese.³ The independent and interdependent items were loosely alternated, such that there was no clear pattern of the two kinds of items.

Last, participants were asked their area of specialization related to Japan, the total number of years that they had spent in Japan, and the total number of years they had spent in North America.

Results and Discussion

Experts' background in Japan. The experts had a diverse range of specializations relevant to Japan, including history, literature, geography, economics, art, political science, and anthropology. They had spent an average of 23.6 years in Japan and an average of 25.6 years in North America. They seem to be in an excellent position to evaluate the two cultures on the items.

Consensus on items. Table 1 summarizes the results for each item. We discovered that university professors are not always willing to be bullied into using only the responses offered them. There were a number of occasions (mostly from 1 individual), shown in the third column, when the respondents answered "no difference" even though this was not an alternative offered them. We calculated consensus agreement for each of the items by dividing the number of responses in the direction of the common view (i.e., Japanese higher for interdependent items and North Americans higher for independent items) by the total of 14 responses ("no difference" responses were counted as 0.5).

The experts were in agreement with the common view for all 23 items, which thus demonstrates face validity for all the items in Singelis's (1994) scale. However, there was some variability regarding the extent of this agreement: The experts viewed some items to better differentiate the cultures than others. We identified the items for which the experts exhibited strong consensus—items for which the perceived cultural differences were unambiguous.

² We used the term North Americans followed by the definition "Canadians and Americans" in parentheses. Mexicans were explicitly excluded from this group.

³ We excluded the item "I feel comfortable using someone's first name soon after I meet them, even when they are much older than I am" because this item is not meaningful in a Japanese context (i.e., even best friends often refer to each other by their last names).

Table 1
Cultural Experts' Evaluation of the Items

Item	No. who chose more characteristic of Japan	No. who chose more characteristic of North America	No. who chose no difference	% consensus
Interdependent				
I have respect for the authority figures with whom I interact.	13	0	1	96
It is important for me to maintain harmony within my group.	14	0	0	100
My happiness depends on the happiness of those around me.	11	2	1	82
I would offer my seat in a bus to my professor.	12	1	1	89
I respect people who are modest about themselves.	14	0	0	100
I will sacrifice my self-interest for the benefit of the group I am in.	13	0	1	96
I often have the feeling that my relationships with others are more important than my own accomplishments.	13	0	1	96
I should take into consideration my parents' advice when making education/career plans.	11	1	2	86
It is important to me to respect decisions made by the group.	13	0	1	96
I will stay in a group if they need me, even when I'm not happy with the group.	14	0	0	100
If my brother or sister fails, I feel responsible.	14	0	0	100
Even when I strongly disagree with group members, I avoid an argument.	14	0	0	100
Independent				
I'd rather say "no" directly, than risk being misunderstood.	2	12	0	86
Speaking up during a class is not a problem for me.	0	14	0	100
Having a lively imagination is important to me.	0	12	2	93
I am comfortable with being singled out for praise or rewards.	0	13	1	96
I am the same person at home that I am at school.	0	13	1	96
Being able to take care of myself is a primary concern for me.	1	12	1	89
I act the same way no matter who I am with.	0	13	1	96
I prefer to be direct and forthright when dealing with people I've just met.	0	14	0	100
I enjoy being unique and different from others in many respects.	0	13	1	96
My personal identity, independent of others, is very important to me.	0	14	0	100
I value being in good health above everything.	2	8	4	71

We selected 95% agreement as the threshold for this criterion (this excludes all items in which at least 1 person held an opinion opposite to the common view), and 16 of the items met or beat this standard. These 16 items were used in the remaining studies as the expert-approved items by which the cultures are seen to differ.

Reconciling Study 1 with the meta-analyses. Our pilot study and Study 1 produce two conflicting sources of evidence. Our literature review of the trait, attitude, and value measures of IC revealed no cultural differences, whereas the cultural experts viewed clear cultural differences between Japan and North America, consistent with the common view. Moreover, as we summarize above, there is consistent support for the common view in a variety of other disciplines as well. How do we reconcile this conflicting evidence? One interpretation, offered by Matsumoto (1999) and Takano and Osaka (1999), is that the psychometric data are correct and the common view is wrong: Japanese are just as individualistic and collectivistic as North Americans. Takano and Osaka maintained that the notion that Japanese are collectivistic derives from inaccurate stereotypes formed during the Meiji Restoration through World War II, when Japanese were pressured to act collectively by unusual circumstances, and that these stereotypes have persisted because of confirmation biases. Another explanation, offered by Matsumoto (1999), is that Japanese society has lost most of its collectivistic heritage through cultural change.

Our position, in contrast, is that given the theoretical problems of different reference groups in comparing cultures on subjective Likert scales, we cannot be confident in the validity of past studies of IC that have used such measures. Hence, we suspect that whereas the cultural experts are likely reasonably accurate in their assessments, the IC comparisons are confounded by the reference-group effect. In Studies 2 and 3, we test empirically whether IC measurements are confounded by the different reference groups of the cultures being compared.

Study 2: The Effects of Manipulating Reference Groups

In none of the past cross-cultural studies of IC were participants provided with any information regarding reference groups by which to evaluate themselves. It seems reasonable to assume, then, that North American college students were evaluating themselves with implicit comparisons with other North American college students, not with East Asian college students, and vice versa for East Asian students. In Study 2, we examined whether explicitly manipulating participants' reference groups would affect their responses to self-report measures of IC. That is, we compared how individuals evaluated themselves on a standard IC measure with how they evaluated themselves when asked to specifically compare themselves with either Japanese or Canadians.

Evaluating oneself in comparison with people of a foreign country makes little sense in the absence of knowledge about the other culture. Thus, we selected people who had experience in two distinct cultures: Specifically, we targeted Canadians who had returned from teaching English in Japan and Japanese exchange students who were living in Canada. Both of these bicultural groups have a considerable amount of knowledge of the other cultural group, such that they can make meaningful assessments of how they compare against people whom they have known in these other cultures.

Method

Participants. Canadians who had been former participants in the Japan Exchange of Teachers program and who had taught English in Japan were approached at a gathering in Vancouver and invited to participate in a study. Participants were of a variety of different ethnic backgrounds, but as we wanted to contrast those of unambiguously Western upbringing, we only included the responses of those of European descent in the analyses. A total of 27 (18 women and 9 men) Euro-Canadians completed a questionnaire in return for a token gift. Japanese students from Ritsumeikan University in Kyoto, who were on an exchange program at the University of British Columbia, were invited to complete a questionnaire in return for extra course credit. A total of 68 did so (44 women, 23 men, and 1 who did not report his or her gender).

Measures. Participants completed three different versions of Singelis's (1994) Independent/Interdependent Self Scale. A no-referent version contained the items in standard form; for example, "I have respect with the authority figures with whom I interact." The second version asked participants to evaluate themselves compared with Japanese; for example, "Compared to most Japanese, I think I have respect with the authority figures with whom I interact." The third version asked participants to evaluate themselves compared with North Americans; for example, "Compared to most North Americans, I think I have respect with the authority figures with whom I interact." Each participant completed all three versions of the questionnaire, which were counterbalanced such that one third completed the no-referent version first, one third completed the Japanese version first, and one third completed the North American version first. At the end of the questionnaire, participants completed some demographics questions, including how long they had spent in Japan and North America. Canadians completed the questionnaire in English, and Japanese in Japanese. The initial English version was translated into Japanese using an extensive translation procedure involving three translators (Heine et al., 2001).

Results and Discussion

Composition of the samples. The two samples did not differ in terms of sex proportions ($\chi^2 < 1$); however, as some studies have found sex differences in IC (Y. Kashima et al., 1995), we included sex as a factor in the analyses. No significant sex effects or Sex \times Culture interactions emerged. The Canadian sample was significantly older ($M = 27.2$) than the Japanese sample ($M = 20.3$), $F(1, 91) = 99.50, p < .001$. Within the Canadian sample (but not within the Japanese one), age was significantly correlated with responses on the no-referent and North American referent independent scales, so we included age as a covariate in all analyses that included these variables, and we report the adjusted means.

Canadians had spent an average of 38.9 months in Japan (with a range of 12 to 78 months), whereas the Japanese students had all been in Canada for 6 months at the time of the study, with a small minority of them having spent additional time in North America prior to participating in the exchange program (total $M = 7.8$

months, ranging from 6 to 36 months). In sum, all participants had a fair degree of exposure to both cultures.

Cultural comparisons of independence and interdependence. All analyses are conducted on the 16 items that the cultural experts from Study 1 agreed unambiguously distinguished North Americans and Japanese. These items provide the strongest test of whether the reference-group effect contaminates participants' responses (the same pattern of results held when all 23 items were included).

We first considered whether the order of the questionnaire had any impact on participants' responses. It did not (all $ps > .20$), and, hence, we collapsed the three orders for all subsequent analyses.

We conducted a three-way analysis of variance (ANOVA) including culture as a between-groups variable, and scale type (Independence vs. Interdependence) and referent version (no referent, cross-cultural referent, or same-cultural referent) as within-group variables. This analysis yielded a highly significant three-way interaction, $Rao R(2, 84) = 36.42, p < .001 (\eta^2 = .46)$, suggesting that the reference groups affected the cultures differently, depending on scale type.

Next we conducted simple effect analyses. Although the participants had exposure to both cultures, the bicultural Canadians had spent most of their lives in Canada, and the bicultural Japanese had spent most of their lives in Japan. Hence, the common view predicts that our Canadian sample should be more independent and less interdependent than our Japanese sample. An ANOVA revealed that Canadians scored significantly higher on the no-referent measure of independence than did Japanese, $F(1, 87) = 11.17, p < .01 (\eta^2 = .16)$, a result that is consistent with the common view (see Table 2). Canadians also scored nominally but not significantly higher on the no-referent measure of interdependence than did Japanese, $F(1, 89) = 2.36, ns$. This result is inconsistent with the common view. A succinct overall test of the interaction between scale type (independence vs. interdependence) and culture. This interaction was marginally significant, $F(1, 88) = 2.90, p < .09 (\eta^2 = .03)$, providing only weak support for the common view.

We repeated these analyses when participants were explicitly comparing themselves with people from the other culture: that is, bicultural Canadians comparing themselves with Japanese and

Table 2
Cross-Cultural Comparisons of Independence and Interdependence

Measure	Canadians	Japanese
No referent		
Independence	3.79 _a	2.99 _b
Interdependence	3.21 _a	2.98 _a
Cross-cultural referent		
Independence	3.81 _a	2.81 _b
Interdependence	2.50 _a	3.24 _b
Same-culture referent		
Independence	3.43 _a	3.34 _a
Interdependence	3.35 _a	2.62 _b

Note. Means in the same column with different subscripts are significantly different at $p < .05$.

bicultural Japanese comparing themselves with Canadians. This is the comparison cross-cultural psychologists have in mind when they discuss cultural differences between Japanese and Canadians. Analyses of independence revealed that Canadians endorsed these items significantly more than did Japanese, $F(1, 89) = 45.89, p < .001$ ($\eta^2 = .34$), a result consistent with the common view. Similarly, analyses of the Interdependence scale demonstrated that Japanese were significantly more interdependent than were Canadians, $F(1, 86) = 18.93, p < .001$ ($\eta^2 = .18$), a result also consistent with the common view. The repeated measures interaction between scale type and culture was highly significant, $F(1, 85) = 38.72, p < .001$ ($\eta^2 = .31$), providing clear support for the common view.

We repeated the analyses once more when participants were explicitly comparing themselves with people from their own culture; that is, bicultural Canadians compared themselves with other Canadians and bicultural Japanese compared themselves with other Japanese. Analyses of independence revealed no difference between Canadians and Japanese, $F(1, 88) < 1$, a finding inconsistent with the common view. Canadians viewed themselves as significantly more interdependent than did Japanese, $F(1, 90) = 24.34, p < .001$ ($\eta^2 = .21$), an effect opposite to that of the common view. The repeated measures interaction between scale format and culture was also significant, $F(1, 88) = 7.54, p < .01$ ($\eta^2 = .08$), but in the direction opposite to that of the common view.

These latter findings might seem odd had the participants not lived in the other culture; however, they are precisely what we would predict of the reference-group effect among bicultural samples. That is, Japanese exchange students who have partially acculturated to the West should be expected to feel relatively high on independence and low on interdependence compared with their compatriots in Japan, and, likewise, Canadian English teachers who have partially acculturated to Japan should feel relatively low on independence and high on interdependence compared with their compatriots in Canada. This pattern also suggests that biculturals maintain reference groups that include members from both of the cultures with whom they have had experience.

Contrasts between different referents. Another way to examine the effect of different reference groups is to compare participants' responses to the no-referent format items with their responses to their cross-cultural referent responses. Repeated measure analyses conducted within each culture reveal that Canadians evaluated themselves similarly on independence ($F < 1$) and as significantly less interdependent, $F(1, 24) = 37.46, p < .001$ ($\eta^2 = .65$) when they compared themselves with Japanese, compared with their no-referent format responses. The interaction between scale type (independence vs. interdependence) and referent (no referent or cross-cultural referent) was significant, $F(1, 21) = 11.51, p < .01$ ($\eta^2 = .35$). Conversely, Japanese evaluated themselves as significantly less independent, $F(1, 64) = 22.23, p < .001$ ($\eta^2 = .26$), and significantly more interdependent, $F(1, 65) = 10.26, p < .01$ ($\eta^2 = .14$), when they compared themselves with Canadians compared to their no-referent format responses. The interaction was significant, $F(1, 66) = 25.42, p < .001$ ($\eta^2 = .28$). Participants were thus evaluating themselves quite differently between these two formats, and in the directions predicted by the reference-group effect.

Repeated measure analyses were also conducted between the no-referent condition and the same-country referent within both cultures. These analyses revealed that Canadians evaluated themselves as significantly less independent, $F(1, 24) = 9.56, p < .01$ ($\eta^2 = .29$), and nominally more interdependent, $F(1, 24) = 2.79, p < .11$ ($\eta^2 = .10$), when they compared themselves explicitly with other Canadians than when they were given no referent. The interaction was significant, $F(1, 25) = 12.27, p < .01$ ($\eta^2 = .33$). Japanese viewed themselves as significantly more independent, $F(1, 63) = 13.63, p < .001$ ($\eta^2 = .18$), and significantly less interdependent, $F(1, 65) = 28.51, p < .001$ ($\eta^2 = .31$), when they explicitly compared themselves with other Japanese than when no reference group was provided. The interaction was significant, $F(1, 65) = 27.70, p < .001$ ($\eta^2 = .30$). That the effects of reference group were of similar magnitude between the no-referent and cross-cultural conditions and the no-referent and same-culture conditions suggests that our samples really are bicultural, existing squarely in between their perceptions of the two cultures. We can think of no theoretical account other than the reference-group effect that would predict this pattern of results.

In sum, Study 2 finds that when participants evaluated themselves with a typical measure of IC (i.e., no-referent condition), there was little support for the common view, replicating past cross-cultural studies of IC. However, when they compared themselves with members of the other culture there was clear support for the common view. Moreover, that participants' evaluations moved in the direction of the common view when participants evaluated themselves in contrast with the other culture, compared with their evaluations when no referent was provided, demonstrates that our participants hold the belief that, on average, they differ from those of the comparison culture precisely in ways predicted by the common view. These results are consistent with the notion that cultural differences between East Asians and North Americans on IC are real and that cultural comparisons using standard measures are confounded by the reference effect.

We selected Japanese exchange students living in Canada and Canadian English teachers who had lived in Japan because we wanted to ensure that participants could make meaningful contrasts with the other culture. It would have been meaningless to ask people to compare themselves with reference groups that they knew nothing about. However, as observed in the evaluations with the same-culture referent, these two samples are likely not representative of their cultures in that Japanese exchange students in Canada are almost certainly more Westernized and Canadians who have lived in Japan are almost certainly more Easternized than their respective compatriots. Note that this lack of representativeness should work to reduce the cultural differences in IC (and reverse the cultural differences when participants compare themselves with those from their home culture). That the effects emerged strongly in the cross-cultural reference-group condition despite the fact that we selected samples that are likely less distinct from each other suggests that the cultural differences between more prototypical cultural exemplars might even be larger.

Participants in Study 2 were asked to make an evaluation with respect to how they perceive members from the other culture. It is possible, as Takano and Osaka (1999) and Matsumoto (1999) have argued, that participants from both cultures view Japanese in terms of inaccurate stereotypes. If this were the case, we would expect that Canadians would feel more independent and less interdependent

dent than their (erroneous) stereotyped images of Japanese. Likewise, it is also conceivable (though perhaps less so) that Japanese have come to internalize the inaccurate stereotypes and thus view themselves as less independent and more interdependent than they perceive Canadians to be or, alternatively, that inaccurate stereotypes about Canadians' independence are possessed by Japanese. It could be argued, then, that Study 2 merely reflects participants' different stereotypes about the cultures rather than true cultural differences regarding IC. The same argument could be made for the judgments of our cultural experts in Study 1. To address this possibility in Study 3, we sought to contrast reference groups without explicitly mentioning them, thus preventing participants from responding on the basis of stereotypes.

Study 3: Within-Culture Versus Between-Cultures Comparisons

The reference-group effect is a problem for cross-cultural comparisons when people from different cultures compare themselves with different standards. The problem should be mitigated if people from two cultural groups habitually consider themselves with respect to a similar standard. One naturally occurring situation in which this can be investigated is in comparisons between people of different cultural backgrounds living in the same country. For example, it is reasonable to assume that European Canadians and Asian Canadians are both likely to evaluate themselves in comparison with other Canadians. Thus, they should be making their self-evaluations with respect to a fairly similar standard (although we imagine that Asian Canadians are less likely to rely as much on European Canadian referents as are European Canadians, and vice versa).

The common view predicts that Asian Canadians should be more interdependent and less independent than European Canadians. Moreover, if different reference groups confound cultural comparisons, we would expect larger differences between Asian Canadians and European Canadians (who share the same referent) than between East Asians living in Asia and North Americans living in Canada (who have different referents). To the extent that the common view is incorrect, however, we would not expect Asian Canadians to differ from European Canadians.

The same argument can be made for cultural variation within Japan. Many Japanese have spent significant time in a Western country and thus have been socialized in a cultural environment that theoretically fosters independence and inhibits interdependence, relative to their domestic compatriots. Thus, if the common view is correct, these returnee Japanese should score higher on independence and lower on interdependence than Japanese who have never been abroad. Moreover, to the extent that both groups of Japanese tend to rely on similar referents (i.e., other Japanese), this predicted difference should not be confounded by the reference-group effect. In contrast, if the common view is incorrect, Japanese who have spent time in the West should be no different than their domestic compatriots.

In sum, we anticipated that cross-national comparisons of Canadians and Japanese would be confounded by reference-group effects and would conceal cultural differences, whereas cross-cultural comparisons within Canada and within Japan would not be so confounded, and thus the cultural differences should hold.

Method

A meta-analysis was carried out on all the Canadian and Japanese participants included in various questionnaire studies conducted by Heine and Lehman that included Singelis's (1994) independence/interdependence measure (excluding the studies from the present article). The participants came from studies that were conducted in either Canada or Japan. Participants were enrolled at the following institutions when the data were collected: Kansai Gaikokugo University ($n = 61$), Ritsumeikan University ($n = 285$), Doshisha University ($n = 83$), and Toyama University ($n = 122$) in Japan, and the University of British Columbia ($n = 1,039$) and Simon Fraser University ($n = 167$) in Canada. In addition, a subset of the Canadian sample consisted of recent college graduates who were about to leave the country to teach English in Japan ($n = 74$).

The samples were segregated into the following groups: The Canadian sample ($n = 1,280$) was segregated into those of European descent, whom we term European Canadians (432 women and 247 men), and those of Asian descent, whom we term Asian Canadians (321 women, 159 men, and 5 who did not report their gender). The remaining participants ($n = 116$) were of a variety of ethnic backgrounds (e.g., mixed ethnicities, African, Latin American, Middle-Eastern, Caribbean) and were excluded from the analyses. One hundred fifty-four of the Asian Canadians were born in Canada, whereas the remaining 331 were born in Asia and had, on average, been in Canada for 5.1 years. The Japanese sample ($n = 551$) was broken down into those who had stated that they had spent some time in a Western country (whom we term returnee Japanese: average time spent was 13 months; 88 women and 36 men) and those who had not (whom we term Japan-bound Japanese; 182 women and 245 men).

Results and Discussion

Comparability of samples. The four samples differed from one another with respect to sex proportions, $\chi^2(3, N = 1,710) = 73.0, p < .001$. Sex was included as a factor for all analyses of scale totals.

An ANOVA revealed significant age differences between the four samples, $F(3, 1702) = 22.67, p < .001$. Post hoc comparisons (Tukey's for unequal sample sizes) revealed that the Japan-bound Japanese were significantly younger ($M = 20.1$) than all of the other groups, and European Canadians ($M = 21.6$) were significantly older than the Asian Canadians ($M = 20.7$). Returnee Japanese fell nonsignificantly between the two Canadian groups ($M = 21.2$). Correlational analyses revealed that age was significantly correlated with the scale total of interdependence for Asian Canadians, and all analyses for interdependence with Asian Canadians thus include age as a covariate.

Comparisons of independence and interdependence. First, we compared European Canadians with Asian Canadians on the two measures. European Canadians scored significantly higher on independence,⁴ $F(1, 1152) = 45.42, p < .001$ ($\eta^2 = .04$), and significantly lower on interdependence, $F(1, 1145) = 21.81, p < .001$ ($\eta^2 = .02$), than Asian Canadians (see Table 3). A repeated measures analysis revealed a significant Culture \times Scale Type interaction, $F(1, 1143) = 49.64, p < .001$ ($\eta^2 = .04$). This provides support for the common view among Canadians.

Analyses of how returnee Japanese compared with Japan-bound Japanese also showed consistent differences. Returnee Japanese

⁴ A main effect for sex emerged in the Canadian sample on independence $F(1, 1152) = 14.01, p < .001$, with men scoring higher ($M = 3.41$) than women ($M = 3.28$). There was no sex effect for Canadian interdependence, nor were there any sex effects for the Japanese sample.

Table 3
Comparisons of Cultural Groups in Study 3

Measure	Within-Canada comparisons		Within-Japan comparisons		Cross-national comparisons	
	European Canadians	Asian Canadians	Returnee Japanese	Japan-bound Japanese	Canadians	Japanese
Independence	3.46 _a	3.23 _b	3.34 _a	3.10 _b	3.37 _a	3.19 _b
Interdependence	3.19 _a	3.34 _b	3.11 _a	3.23 _b	3.29 _a	3.15 _b

Note. Means in the same columns with different subscripts are significantly different at $p < .05$.

scored higher on independence, $F(1, 546) = 16.44, p < .001$ ($\eta^2 = .03$), and lower on interdependence, $F(1, 541) = 4.64, p < .04$ ($\eta^2 = .01$), than did Japan-bound Japanese. The Culture \times Scale Type interaction was also significant, $F(1, 540) = 20.86, p < .001$ ($\eta^2 = .04$). This also provides support for the common view among Japanese.

We then analyzed how the two nations compared on the two measures. We conducted an ANOVA on the composite samples of Canadians compared with Japanese, collapsing across the two cultural groups within each country. These analyses revealed that Canadians were more independent than Japanese, $F(1, 1702) = 41.22, p < .001$ ($\eta^2 = .02$), supporting the common view, and that Canadians were also more interdependent than Japanese, $F(1, 1688) = 31.00, p < .001$ ($\eta^2 = .02$), a result inconsistent with the common view. The Culture \times Scale Type interaction was not significant ($F < 1$) despite the very large sample, thus again failing to support the common view. These analyses provides further evidence that reference-group effects confound cross-national comparisons.

What is striking is that the obtained cultural differences in IC were more pronounced within countries, even though these groups theoretically should differ less from each other than across countries. That is, the common view predicts that Canadians should differ more from Japanese than should European Canadians from Asian Canadians or returnee Japanese from Japan-bound Japanese. That the opposite is true is consistent with the notion that reference-group effects confound cross-national comparisons more so than within-country comparisons. These findings cannot be explained by the notion that participants are embracing inaccurate stereotypes of the two cultural groups, nor are we aware of any social or cultural psychological theories that can account for this pattern other than reference-group effects.

General Discussion

The cultural differences between East Asians and North Americans that are assumed to exist by cultural psychological theory, the so-called common view, are not consistently apparent in cross-cultural comparisons of attitude, trait, and value measures (Matsumoto, 1999; Oyserman et al., 2002; Peng et al., 1997; Takano & Osaka, 1999). Yet there is considerable consensus among cultural experts from a variety of fields that the common view is accurate. There is conflict between these different sources of data that is in need of resolution.

Resolution of the conflict requires us to consider shortcomings with both of the approaches: At least one of these sources of data must be inaccurate. One interpretation that has been offered is that

the common view is inaccurate and that it hinges on outdated stereotypes (Matsumoto, 1999; Takano & Osaka, 1999). However, the only evidence in favor of this interpretation is the source of the conflict itself: namely, the failure of past studies of cross-cultural comparisons of attitude, trait, and value measures to distinguish between the two cultures. There have been no additional empirical demonstrations to support the notion that the widely shared common view is inaccurate.⁵

Our interpretation is that the results from cross-cultural comparisons of means from attitude, trait, and value measures are inaccurate. People from different cultures adopt different standards when evaluating themselves on subjective Likert scales. Comparing measures with subjective Likert response options conceals the very cultural differences that confound the comparisons with the reference-group effect. This is problematic.

There is much evidence in support of the common view. As there are a variety of ways to evaluate any question and the methodologies differ across disciplines, it is useful to look at other sources of evidence in support of it. First, there is consensus for the common view among many different fields using their own methodologies. The striking agreement among the cultural experts in Study 1 demonstrates that the common view is indeed widely shared. Second, Studies 2 and 3 reveal that when efforts are made to control for the reference-group effect, the data from attitude and

⁵ In addition to 11 studies that compared the two cultures on IC measures, Takano and Osaka (1999) also highlighted two pairs of behavioral studies that appear to challenge the common view. One of these was the Japanese literature on conformity using the Asch (1956) paradigm, and the other was Yamagishi's (1988) cross-cultural research on the free-rider problem. The conformity research reveals that Japanese do not conform at a particularly high level when in a group of strangers, whereas there is evidence that they do more so when with their in-group members (a study measuring conformity among Japanese in-groups was the second highest out of 133 samples reviewed by Bond & Smith, 1996). In contrast, in-group status seems to have relatively little effect on American conformity (see Bond & Smith, 1996). Similarly, Yamagishi's research (Yamagishi, 1988; Yamagishi & Yamagishi, 1994) suggests that Japanese show little cooperation with strangers, although they do appear to have high levels of trust toward their in-group members (Yamagishi & Yamagishi, 1994). These studies provide evidence that is very much in support of the notion that Japanese are more collectivistic than North Americans. The interdependent self is clearly not interdependent with everyone but rather only with self-defining in-group relationships (also see Iyengar et al., 1999). Thus, we disagree with Takano and Osaka (1999) that these studies represent a challenge to the common view.

trait items are also consistent with the common view (the same can be said for value measures; Peng et al., 1997). Last, cultural psychologists, using methodologies other than comparisons of means of IC scales, have produced much evidence in support of the common view: For example, compared with North Americans, East Asians exhibit more flexible selves, (Kanagawa, Cross, & Markus, 2001; Suh, 2001), are more likely to make situational attributions (Choi, Nisbett, & Norenzayan, 1999; Morris & Peng, 1994), are more likely to show evidence of self-criticism (Heine et al., 1999; Kitayama, Markus, Matsumoto & Norasakkunkit, 1997), show less pronounced motivations for uniqueness, (Kim & Markus, 1999), experience more interdependent emotions (Kitayama, Markus, & Kurokawa, 2000), show less attitude-behavior consistency (Heine & Lehman, 1997b; Y. Kashima, Siegal, Tanaka, & Kashima, 1992), seek greater financial risk (Hsee & Weber, 1999), differ in their perceptions of romantic love (Dion & Dion, 1993), are more likely to adopt the third-person perspective in their memories (Cohen & Gunz, 2002), perceive more distinct boundaries between the in-group and out-group (Buchan, Croson, & Dawes, 1999; Iyengar, Lepper, & Ross, 1999; cf. Yamagishi & Yamagishi, 1994), show more evidence for experiencing agency through adjusting or secondary control (Morling, Kitayama, & Miyamoto, 2002; Weisz, Rothbaum, & Blackburn, 1984), view fulfillment of roles as a more important basis for subjective well-being (Suh et al., 1998), show more evidence of preventive and self-improving motivations (Heine et al., 2001; Lee, Aaker, & Gardner, 2000), and are more likely to demonstrate holistic reasoning styles (Nisbett, Peng, Choi, & Norenzayan, 2001). All of these cultural differences have been attributed, at least in part, to differences related to IC. Indeed, the only psychological evidence that consistently conflicts with the common view is based on subjective Likert measures of IC. We conclude that the common view is alive and well.

It is important to note that, theoretically, the reference-group effect is a problem only when mean scores from different groups with different referents are contrasted. It does not render assessment with subjective Likert scales problematic for identifying variability within a group in which common referents are held. Indeed, one strategy for interpreting cultural differences has been to unpack culture by identifying differences within cultures that correlate with the dimension under question (e.g., Bond, 1994; Heine et al., 2001; Kwan, Bond, & Singelis, 1997; Singelis, Bond, Lai, & Sharkey, 1999), a strategy that we believe is valid and important. For example, if people who score higher on independence are more likely to self-enhance within cultures, we might be able to make sense of cultural differences in self-enhancement by considering cultural differences in independence (e.g., Heine & Renshaw, 2002). The existence of within-culture variability is important in aiding the identification of the underlying psychological mechanisms. The use of subjective Likert scales is most valid for identifying differences within rather than between groups.

Some might question our use of cultural experts as a standard for validity. Surely cultural experts also have their own biases, and these biases might be shared among members of a discipline, leading them to exaggerate the extent of presumed cultural differences. However, the experts' ratings do at least provide a criterion, and in that regard we see them as an improvement over studies that do not provide any validity criteria. Moreover, the validity of our validity criterion is corroborated in that the manipulations in

Studies 2 and 3 that controlled reference-group effects all led participants to respond more in line with the experts' opinions (also see Peng et al., 1997). It is difficult to reconcile this pattern if the experts were not a reasonably valid standard. It is possible that there are other good validity standards that can be used in cross-cultural comparisons, and we think the field will be advanced if other validity criteria can be developed.

We have focused on cultural comparisons of IC, as this is the most studied construct across cultures (in particular, between East Asian and North American cultures), but the conclusions generalize to other constructs as well. For example, a series of studies revealed that, whereas subjective Likert scale comparisons of malleable views of self do not reveal cultural differences (e.g., Chiu et al., 1997), those correcting for reference effects do (Heine et al., 2001). It is nonetheless possible that reference-group effects are more problematic for some kinds of items than others. For example, cross-cultural comparisons of self-esteem between Japanese and North Americans using a wide variety of different methodologies reveal pronounced cultural differences (Heine et al., 1999), even though some of these methodologies also use subjective Likert scales (e.g., Heine & Lehman, in press). Consistent cultural differences are found both in studies that measure well-being with subjective Likert scales (Diener, Diener, & Diener, 1995) and in those that use experimental methods (e.g., Oishi, Wyer, & Colcombe, 2000). Why do the different reference groups of the cultures not conceal cultural differences with these measures and items? One possibility is that participants rely less on social comparison when responding to these kinds of items. Questions such as whether individuals feel satisfied with their life or feel confident about their abilities might rely more on introspection and comparison with internal standards than on implicit comparisons with consensually shared standards, which thus might mute the effects of different referents. Other items that should be less affected by reference groups are those that measure concrete behaviors, as the responses indicate whether or how frequently these have occurred rather than their relative strength inferred through social comparison (e.g., Biernat & Manis, 1994). Likewise, it seems that Likert scale options that are more concrete, such as *at least twice a day* rather than *strongly agree*, are better protected, as the responses are tethered to a more objective referent. Hence, some kinds of items seem to be less influenced by reference-group effects, and more research is necessary to determine the kinds of items that are most problematic. In general, our recommendation is to assume that cross-cultural comparisons of means with subjective Likert scales are confounded by reference-group effects unless similar results are obtained with other measures.

The reference-group effect is not an issue only for cultural comparisons but for any comparison of subjective Likert scale response between groups that possess different referents. The most obvious example is comparisons between the sexes (e.g., Biernat et al., 1991), but it might also confound comparisons between generations (e.g., parents vs. children) or organizations (e.g., different schools or companies). Indeed, it is hard to make sense of the finding that 35-year olds rate themselves to be as healthy as do 80-year olds (Nillson et al., 1997) without invoking the reference-group effect. It seems to us, though, that the reference-group effect is especially problematic for cultures. People regularly interact with members of the opposite sex, different generations, or differ-

ent organizations, so these other referents are not beyond their consideration. However, most people's interactions tend to be with members of their own culture, and they may have little contact with the kinds of cultural groups that are contrasted in cross-cultural psychological studies. This tendency to exist within a single cultural world leads culture to be largely invisible to people. People conflate what is true of their culture with what they assume to be true of human nature, thus projecting their own ways of thinking onto the rest of the world. It is telling that many of the people who have become cross-cultural psychologists are those who have led bicultural lives and were thus in a position to notice the cultural differences.

What Can We Do?

What strategies can be adopted to avoid the confounds associated with the different referents that different cultures possess? There is unlikely to be a single magic bullet; however, there are a variety of methodologies that seem less compromised by the reference-group effect.

One strategy to avoid reference-group effects is to avoid measuring culture through individual-level responses. To the extent that culture represents shared beliefs and norms, it might be more appropriate to investigate culture by examining cultural-level measures. A recent investigation of regional IC differences within the 50 states of the United States by Vandellos and Cohen (1999) nicely demonstrates this approach. Some of their measures included cultural-level variables such as the divorce rate, percentage of nuclear families, voting patterns, and different laws. Kim and Markus (1999) and Han and Shavit (1994) used a similar approach in examining the kinds of cultural messages evident in magazine advertisements, a reflection of the shared ideas of the cultures. This approach has the distinct advantage of revealing the culture directly rather than attempting to view it through the lens of the individual and avoids various methodological problems associated with reference-group effects, response sets, and translation problems. Indeed, the pattern of IC that emerged through these studies was consistent with the theoretical literature. However, this approach is not without its shortcomings. Whereas such comparisons might be straightforward within a country, comparisons across countries with different governments, legal systems, infrastructures, and so forth render such comparisons more problematic. And it is important to note that cultural-level measures do not allow researchers to unpackage cultural differences at the psychological level.

Another strategy is to have people compare themselves with an arithmetic standard. For example, some comparisons of self-enhancement across cultures ask people to compare themselves with the average person in their class or to estimate the percentage of people in their class who are more extreme than they are on a dimension (e.g., Heine & Lehman, 1997a; Markus & Kitayama, 1991a). As these standards are derived mathematically from within each sample, they serve to equate the cultures on the question under study, namely, how people view themselves in comparison with their classmates. However, this methodology only seems useful in investigations of how people view themselves in contrast to a particular standard.

A reliable strategy used in many recent experiments comparing cultures (e.g., Lee et al., 2000; Peng, & Nisbett, 1999; Sanchez-

Burks, Nisbett, & Ybarra, 2000) is to contrast responses across more than one condition within each culture. As the comparisons concern patterns between conditions within each culture, any effects of reference groups or other confounds should be controlled. For example, Iyengar and Lepper (1999) investigated intrinsic motivation among children in an experimental task. Although they did not find cultural differences in persistence when children chose the task themselves or when the experimenter assigned the children the task, the results differed greatly when children believed that their mother had assigned them the task. Had they only looked at the mother condition, alternative explanations could be offered, such as that the task was more interesting for children of one culture than of the other. That the cultural difference only emerged in the mother condition, however, suggests that there is something about children's relationships with their mother that affected the results. Cultural differences in the pattern across conditions is compelling evidence for cultural influences on the psychological process under study. In general, we think this is an excellent strategy for investigating culture; however, one serious drawback is that this experimental method is not particularly useful for identifying individual differences within cultures.

Another strategy is to determine how different cultures respond to information that is either consistent or inconsistent with implicit theories that are hypothesized to be shared by members of a culture (e.g., Heine et al., 2001). The reasoning behind this approach is that if instructions lead to different behavior compared with a control group, the instructions probably conveyed new information not part of widely shared lay theories of the culture. In contrast, if the instructions have no impact on behavior, one could assume that they are conveying no new information: The instructions are redundant with the lay theory. Heine et al. (2001) contrasted Japanese and American persistence behavior after failure of a no-instructions group with that of groups that received instructions highlighting the utility of effort or the lack of utility of effort for an experimental task. Results indicated that Americans in the high-utility condition persisted longer than those in a control condition but those in a low-utility condition did not differ from the control. In contrast, Japanese in the high-utility condition did not differ from those in a control condition, whereas those in a low-utility condition persisted less than those in a control condition. Thus, it appears that Americans held the lay belief that the particular experimental task was one that measured fixed abilities, and efforts would have little impact on performance, whereas Japanese held the belief that efforts would enhance performance. This methodology might be of use for investigating the presence of other lay theories widely held within cultures; however, it is unable to address individual differences within cultures.

Some kinds of dependent measures are immune to reference-group effects. When behavioral measures are used, such as when people measure aggression by playing chicken in the hallway (Cohen, Nisbett, Bowdle, & Schwarz, 1996) or when motivation is measured by an unobtrusive observer timing persistence (Heine et al., 2001; Iyengar & Lepper, 1999), many self-report-based confounds are minimized. Physiological measures such as testosterone levels (Cohen et al., 1996), heart rate, or skin conductance (Levenson, Ekman, Heider, & Friesen, 1992) are especially resistant to reference-group effects and other confounds. The benefits of these strategies, however, are somewhat offset by the considerable time and cost involved in using them as well as by the difficulty of

determining appropriate behavioral and physiological indices of constructs such as IC.

A final strategy is to use a forced-choice framework that includes items with concrete, objective response options. This method has been shown to enhance the validity of comparisons across groups in much past research (e.g., Biernat & Manis, 1994; Heine et al., 2001, Study 4; Peng et al., 1997). However, this approach, too, has its downsides. Objective responses tend to be more specific in contrast with the breadth that attitude, trait, or value statements encompass, thus limiting the generalizability of the findings. Nonetheless, we feel these costs tend to be outweighed by the benefits of concrete response type questions in comparison with subjective Likert measures when researchers are contrasting cultures.

This is surely not an exhaustive list of strategies for cross-cultural comparisons that protect against reference-group effects, and we believe that the field will benefit by the development of new methodologies that reduce self-report confounds. Each of the approaches listed above has weaknesses, and none is appropriate in all situations. So which is the best method for studying cultural similarities and differences? The most successful approach is likely one that combines many of these and uses multiple methodologies (Triandis et al., 1990). To the extent that results converge with multiple measures, we can be more confident that cultural similarities or differences are not due to artifacts. The multiple methodologies provided by Nisbett and Cohen (1996) in their study of the culture of honor in the American South represent a prototypical example of this approach. Reliance on one single method, particularly one confounded by reference-group effects, yields results of dubious validity.

References

- Asch, S. (1956). Studies of independence and conformity. A minority of one against a unanimous majority. *Psychological Monographs*, 70(9, Whole No. 416).
- Bakan, D. (1966). *The duality of human existence*. Chicago: Rand McNally.
- Baldwin, M. W., & Holmes, J. G. (1987). Salient private audiences and awareness of the self. *Journal of Personality and Social Psychology*, 52, 1087–1098.
- Befu, H. (1980). A critique of the group model of Japanese society. *Social Analysis*, 5(6), 29–43.
- Biernat, M., & Billings, L. S. (in press). Standards, expectancies, and social comparison. In N. Schwarz & A. Tesser (Eds.), *Blackwell handbook of social psychology, Volume 1: Intrapersonal processes*. London: Blackwell.
- Biernat, M., & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, 72, 544–557.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66, 5–20.
- Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, 60, 485–499.
- Bond, M. H. (1988). Finding universal dimensions of individual variation in multicultural studies of values: The Rokeach and Chinese value surveys. *Journal of Personality and Social Psychology*, 55, 1009–1015.
- Bond, M. H. (1994). Emotions and their expression in Chinese culture. *Journal of Nonverbal Behavior*, 17, 245–262.
- Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, 119, 111–137.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185–216.
- Buchan, N., Croson, R., & Dawes, R. (1999, November). *Who's with me? The role of group boundaries on trust and reciprocity: A cross-cultural study*. Paper presented at the meeting of the Society for Judgment and Decision Making, Los Angeles, CA.
- Buruma, I. (1984). *A Japanese mirror: Heroes and villains of Japanese culture*. Middlesex, England: Penguin Books.
- Campbell, D. T., Lewis, N. A., & Hunt, W. A. (1958). Context effects with judgmental language that is absolute, extensive, and extra-experimentally anchored. *Journal of Experimental Psychology*, 55, 220–228.
- Chen, C., Lee, S.-Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6, 170–175.
- Chiu, C., Hong, Y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology*, 73, 19–30.
- Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin*, 125, 47–63.
- Christopher, R. C. (1983). *The Japanese mind*. New York: Fawcett Columbine.
- Cohen, D., & Gunz, A. (2002). As seen by the others: The self from the “outside in” and the “inside out” in the memories and emotional perceptions of Easterners and Westerners. *Psychological Science*, 13, 55–59.
- Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An “experimental ethnography.” *Journal of Personality and Social Psychology*, 70, 945–960.
- Collins, R. J. (1992). *Japan-think, Ameri-think*. New York: Penguin Books.
- Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin*, 119, 51–69.
- Diener, E., Diener, M., & Diener, C. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology*, 69, 851–864.
- Dion, K. K., & Dion, K. L. (1993). Individualistic and collectivistic perspectives on gender and the cultural context of love and intimacy. *Journal of Social Issues*, 49, 53–69.
- Doi, T. (1971). *The anatomy of dependence*. Tokyo: Kodansha.
- Feiler, B. S. (1991). *Learning to bow*. Boston: Houghton-Mifflin.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140.
- Greenfield, P. M. (1997). Culture as process: Empirical methods for cultural psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (Vol. 1, pp. 301–346). Boston: Allyn & Bacon.
- Hamaguchi, E. (1985). A contextual model of the Japanese: Toward a methodological innovation in Japan studies. *Journal of Japanese Studies*, 11, 289–321.
- Han, S., & Shavit, S. (1994). Persuasion and culture: Advertising appeals in individualistic and collectivist societies. *Journal of Experimental Social Psychology*, 30, 326–350.
- Heine, S. J. (2001). Self as cultural product: An examination of East Asian and North American selves. *Journal of Personality*, 69, 881–906.
- Heine, S. J., Kitayama, S., Lehman, D. R., Takata, T., Ide, E., Leung, C., & Matsumoto, H. (2001). Divergent consequences of success and failure in Japan and North America: An investigation of self-improving motivations and malleable selves. *Journal of Personality and Social Psychology*, 81, 599–615.
- Heine, S. J., & Lehman, D. R. (1997a). The cultural construction of self-enhancement: An examination of group-serving biases. *Journal of Personality and Social Psychology*, 72, 1268–1283.

- Heine, S. J., & Lehman, D. R. (1997b). Culture, dissonance, and self-affirmation. *Personality and Social Psychology Bulletin*, 23, 389–400.
- Heine, S. J., & Lehman, D. R. (in press). Move the body, change the self: Acculturative effects on the self-concept. In M. Schaller & C. Crandall (Eds.), *Psychological foundations of culture*. Mahwah, NJ: Erlbaum.
- Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106, 766–794.
- Heine, S. J., & Renshaw, K. (2002). Interjudge agreement, self-enhancement, and liking: Cross-cultural divergences. *Personality and Social Psychology Bulletin*, 28, 442–451.
- Heine, S. J., Takata, T., & Lehman, D. R. (2000). Beyond self-presentation: Evidence for self-criticism among Japanese. *Personality and Social Psychology Bulletin*, 26, 71–78.
- Hendry, J. (1987). *Understanding Japanese society*. New York: Crook-Helm.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- Hsee, C., & Weber, E. U. (1999). Cross-national differences in risk preference and lay predictions. *Journal of Behavioral Decision Making*, 12, 165–179.
- Hyman, H. (1942). The psychology of subjective status. *Psychological Bulletin*, 39, 473–474.
- Ichiki, K. (1985). "Ijime" no nichibei hikaku [A comparison between Japan and the U. S. concerning "bullying"]. *Kyouiku to Igaku*, 33, 919–924.
- Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*, 76, 349–366.
- Iyengar, S. S., Lepper, M. R., & Ross, L. (1999). Independence from whom? Interdependence with whom? Cultural perspectives on ingroups versus outgroups. In D. Miller & D. Prentice (Eds.), *Cultural divides: Understanding and overcoming group conflict* (pp. 273–301). New York: Sage.
- Johnson, F. A. (1993). *Dependency and Japanese socialization*. New York: New York University Press.
- Kagitcibasi, C. (1994). A critical appraisal of individualism and collectivism: Toward a new formulation. In U. Kim, H. C. Triandis, C. Kagitcibasi, S.-C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (pp. 52–65). Thousand Oaks, CA: Sage.
- Kanagawa, C., Cross, S. E., & Markus, H. R. (2001). "Who am I?": The cultural psychology of the conceptual self. *Personality and Social Psychology Bulletin*, 27, 90–103.
- Kashima, E. S., & Kashima, Y. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-Cultural Psychology* 29, 461–486.
- Kashima, Y., Siegal, M., Tanaka, K., & Kashima, E. S. (1992). Do people believe behaviors are consistent with attitudes? Towards a cultural psychology of attribution processes. *British Journal of Social Psychology*, 31, 111–124.
- Kashima, Y., Yamaguchi, S., Kim, U., Choi, S. C., Gelfand, M. J., & Yuki, M. (1995). Culture, gender, and self: A perspective from individualism-collectivism research. *Journal of Personality and Social Psychology*, 69, 925–937.
- Kim, H., & Markus, H. R. (1999). Deviance or uniqueness, harmony or conformity? A cultural analysis. *Journal of Personality and Social Psychology*, 77, 785–800.
- Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, emotion, and well-being: Good feelings in Japan and the United States. *Cognition and Emotion*, 14, 93–124.
- Kitayama, S., Markus, H. R., Matsumoto, H., & Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality and Social Psychology*, 72, 1245–1267.
- Kiuchi, A. (1995). Construction of a scale for independent and interdependent construals of self and its reliability and validity. *Japanese Journal of Psychology*, 66(2), 100–106.
- Kondo, D. K. (1990). *Crafting selves: Power, gender, and discourses of identity in a Japanese workplace*. Chicago: University of Chicago Press.
- Krantz, D. L., & Campbell, D. T. (1961). Separating perceptual and linguistic effects of context shifts upon absolute judgments. *Journal of Experimental Psychology*, 62, 35–42.
- Kwan, V. S. Y., Bond, M. H., & Singelis, T. M. (1997). Pancultural explanations for life satisfaction: Adding relationship harmony to self-esteem. *Journal of Personality and Social Psychology*, 73, 1038–1051.
- Latane, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts.
- Lebra, T. S. (1976). *Japanese patterns of behavior*. Honolulu: University of Hawaii Press.
- Lee, A. Y., Aaker, J. L., & Gardner, W. L. (2000). The pleasures and pains of distinct self-construals: The role of interdependence in regulatory focus. *Journal of Personality and Social Psychology*, 78, 1122–1134.
- Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra. *Journal of Personality and Social Psychology*, 62, 972–988.
- Lewis, C. C. (1995). *Educating hearts and minds*. New York: Cambridge University Press.
- Manis, M. (1967). Context effects in communication. *Journal of Personality and Social Psychology*, 5, 325–334.
- Markus, H. R., & Kitayama, S. (1991a). Cultural variation in the self-concept. In G. R. Goethals & J. Strauss (Eds.), *Multidisciplinary perspectives on the self* (pp. 18–48). New York: Springer-Verlag.
- Markus, H. R., & Kitayama, S. (1991b). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224–253.
- Maslow, A. (1943). A theory of human motivation. *Psychological Review*, 50, 370–396.
- Matsumoto, D. (1999). Culture and self: An empirical assessment of Markus and Kitayama's theory of independent and interdependent self-construal. *Asian Journal of Social Psychology*, 2, 289–310.
- Miyamoto, M. (1994). *Straitjacket society*. Tokyo: Kodansha.
- Morling, B., Kitayama, S., & Miyamoto, Y. (2002). Cultural practices emphasize influence in the United States and adjustment in Japan. *Personality and Social Psychology Bulletin*, 28, 311–323.
- Morris, M., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, 67, 949–971.
- Mouer, R. E., & Sugimoto, Y. (1986). *Images of Japanese society: A study in the social construction of reality*. London: Routledge & Kegan Paul.
- Nakamura, H. (1964). *Ways of thinking of Eastern peoples*. Honolulu: University of Hawaii Press.
- Nakane, C. (1970). *Japanese society*. Los Angeles: University of California Press.
- Nilsson, L.-G., Backman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., et al. (1997). The Betula prospective cohort study: Memory, health, and aging. (Data from first test session). *Aging, Neuropsychology, and Cognition*, 4, 1–32.
- Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the south*. Boulder, CO: Westview Press.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review*, 108, 291–310.
- Oishi, S., Wyer, R. S. Jr., & Colcombe, S. (2000). Cultural variation in the use of current life satisfaction to predict the future. *Journal of Personality and Social Psychology*, 78, 434–445.
- Oyserman, D., Coon, H. M., & Kummelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128, 3–72.

- Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, *54*, 741–754.
- Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods*, *2*, 329–344.
- Reischauer, E. O. (1988). *The Japanese today: Change and continuity*. Tokyo: Tuttle.
- Rhee, E., Uleman, J. S., & Lee, H. K. (1996). Variations in collectivism and individualism by ingroup and culture: Confirmatory factor analysis. *Journal of Personality and Social Psychology*, *71*, 1037–1054.
- Roland, A. (1988). *In search of self in India and Japan*. Princeton, NJ: Princeton University Press.
- Rosenberger, N. R. (1992). *Japanese sense of self*. Cambridge, England: Cambridge University Press.
- Sanchez-Burks, J., Nisbett, R. E., & Ybarra, O. (2000). Cultural styles, relational schemas and prejudice against outgroups. *Journal of Personality and Social Psychology*, *79*, 174–189.
- Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, S.-C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (pp. 85–119). Thousand Oaks, CA: Sage.
- Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, *58*, 878–891.
- Schwartz, S. H., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, *26*, 92–116.
- Sherif, M. A. (1936). *The psychology of social norms*. New York: Harper.
- Shweder, R. A., Goodnow, J., Hatano, G., LeVine, R. A., Markus, H., & Miller, P. (1998). The cultural psychology of development: One mind, many mentalities. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology* (Vol. 1, pp. 865–937). New York: Wiley.
- Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, *20*, 580–591.
- Singelis, T. M., Bond, M. H., Lai, S. Y., & Sharkey, W. F. (1999). Unpackaging culture's influence on self-esteem and embarrassability: The role of self-construals. *Journal of Cross-Cultural Psychology*, *30*, 315–331.
- Suh, E. M. (2001). *Culture, identity consistency, and subjective well-being*. Manuscript submitted for publication, University of California, Irvine.
- Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology*, *74*, 482–493.
- Takano, Y., & Osaka, E. (1999). An unsupported common view: Comparing Japan and the U. S. on individualism/collectivism. *Asian Journal of Social Psychology*, *2*, 311–341.
- Takata, T. (1999). Nihon bunka ni okeru sougou dokuritsusei, sougou kyouchousei no hattatsu katei: Hikaku bunkateki oudanteki shiryō ni yoru jishshouteki kentō [Developmental process of independent and interdependent self-construal in Japanese culture: Cross-cultural and cross-sectional analyses]. *Japanese Journal of Educational Psychology*, *47*, 480–489.
- Tobin, J. J., Wu, D. Y. H., & Davidson, D. (1989). *Preschool in three cultures*. New Haven, CT: Yale University Press.
- Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, *96*, 506–520.
- Triandis, H. C. (1996). The psychological measurement of cultural syndromes. *American Psychologist*, *51*, 407–415.
- Triandis, H. C., McCusker, C., & Hui, H. (1990). Multimethod probes of individualism and collectivism. *Journal of Personality and Social Psychology*, *59*, 1006–1020.
- Vandello, J. A., & Cohen, D. (1999). Patterns of individualism and collectivism across the United States. *Journal of Personality and Social Psychology*, *77*, 279–292.
- Vogel, E. F. (1979). *Japan as number one: Lessons for America*. Cambridge, MA: Harvard University Press.
- Volkman, J. (1951). Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper.
- Weisz, J. R., Rothbaum, F. M., & Blackburn, T. C. (1984). Standing out and standing in: The psychology of control in America and Japan. *American Psychologist*, *39*, 955–969.
- Whiting, R. (1990). *You gotta have wa*. New York: Vintage Departures.
- Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, *90*, 245–271.
- Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin*, *106*, 231–248.
- Yamagishi, T. (1988). Exit from the group as an individualistic solution to the free rider problem in the United States and Japan. *Journal of Experimental Social Psychology*, *24*, 530–542.
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*, 9–66.

Received December 15, 2000

Revision received October 5, 2001

Accepted October 13, 2001 ■