Understanding Social Psychology Across Cultures

Engaging With Others in a Changing World

2ND EDITION

Peter B. Smith, Ronald Fischer, Vivian L. Vignoles
and Michael Harris Bond

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

4

How to do Cross-Cultural Psychology

I cannot rest from travel …
For always roaming with a hungry heart
Much have I seen and known;
cities of men and manners, climates, councils, governments …
To follow knowledge like a sinking star,
Beyond the utmost bound of human thought…

(from *Ulysses*, by Alfred Lord Tennyson)

Thus far we have outlined some of the intellectual challenges and major insights into how cultures have emerged and how they differ today in their values and beliefs. Before we continue with our quest to chart the cultural side of social psychology, we need to pause and consider some of the practical issues involved in doing and interpreting cross-cultural research, and in researching the cultural diversities that exist on our planet. Take a look at Everyday Life Box 4.1 below.

Everyday Life Box 4.1 presents distressing events occurring among a tribe who had been displaced from their traditional land in a remote area of Eastern Africa. Imagine that you are a researcher and you are encountering a cultural group in which people routinely behave like the Ik. How could you make sense of these incidents as a psychologist? In this chapter we examine the question of how we can conduct research across cultures. What methods can you use and what are the particular challenges or obstacles that await you if you venture into the field? Throughout the chapter we will identify a series of guidelines which summarise how best to face up to these challenges.

Everyday Life Box 4.1

Anthropologists have sought to create detailed pictures of everyday life in a very wide range of social contexts. In a particularly vivid study, Turnbull (1972) documented the progressive effects of very severe famine among the Ik tribal group in the mountains of Uganda. After the oldest people had died, the children were next at risk. In the face of acute hunger, Turnbull noted the way that the other children turned against a child named Adupa, snatching whatever food she was able to obtain from her as she became weaker, and even teasing her by giving her food and then taking it away from her again. As days passed, Turnbull was unable to bear watching this process and compromised his observer status by starting to give her food, but this only prolonged her agony. When she turned to her parents they rebuffed her. Eventually her parents placed her in a hut and sealed it up until she had died, thus raising the survival chances of other children. Turnbull observed that the parents preserved good relations with their neighbours by taking Adupa's remains a good distance away and preventing wild animals from scattering body parts onto their land.

One issue in particular that will demand our attention is how to balance the typical emphasis on the individual in psychological research with the fact that cultural processes do not exist in isolation, but exist within collectives and are shared between individuals. This inter-individual nature of culture requires analyses at a collective level. In Chapter 2, we saw that the landmark studies in cross-cultural psychology have been made at the nation level. Now we must consider how we can use these insights to understand the experiences and behaviours of individuals. We will present core issues, the debates around these and their implications with examples drawn from real and illustrative research, returning to the plight of the Ik mountain people at several points in the chapter.

ETHNOGRAPHIC FIELDWORK

The first option available to cross-cultural researchers going into the field was to use ethnographic research. Travelers for millennia have recorded and reported the customs of the peoples they visited. Marco Polo's famous account of travels to what is today India and China is one of the first examples in the Western world. Marco Polo's near contemporary, the Muslim Ibn Batutta, provides an equally fascinating account from a non-Western viewpoint of a much longer and further journey. These explorers collected extensive but casual observations while working and living among alien cultural groups, often describing the exotic and bizarre events that accentuated the differences between the Venetian and Berber societies that they hailed from and the strange new worlds that they encountered.

The Polish anthropologist, Bronislaw Malinowski (1922), challenged this idiosyncratic method in his classic study *Argonauts of the Western Pacific*, in which he first outlined the scientific method of ethnographic fieldwork for the new discipline of anthropology. He emphasised that a researcher must have the scientific goal of comprehending the complete society rather than reporting selected facts of living that are salient to the foreign observer ('this goal is, briefly,

to grasp the native's point of view, his relation to life, to realize *his* vision of *his* world', p. 25), to live and participate completely in the culture of interest (a process that we now call 'cultural immersion') and to apply scientific methods of collecting, reporting and interpreting evidence:

> In Ethnography, the writer is his own chronicler and the historian at the same time, while his sources are no doubt easily accessible, but also supremely elusive and complex; they are not embodied in fixed, material documents, but in the behavior and in the memory of living men. (1922: 3)

As you can imagine, living with the Ik while they struggled with a constant state of famine was a major physical and psychological struggle for Turnbull. Staying on location, remaining professional and maintaining his ethical sensibilities were major issues for him. His study has been lauded as showing humanity and objectivity even in the face of great suffering and societal collapse, while not offering judgments about the lives and culture of others, even if what they do seems inhumane and cruel to us. This is a reminder of the conclusions that we drew in Chapter 1, that researchers should carefully consider alternative cultural viewpoints when interpreting their research findings.

Guideline 1: Researchers should recognise the cultural contingencies of their own values and beliefs when conducting research and should guard against evaluating other cultures against the criteria of their own cultural understanding.

Ethnographic methods have been the hallmark of social anthropology, cultural psychology, and some approaches to indigenous psychology for nearly a century. The startling account of Ik society is based on exactly these ethnographic methods. Turnbull was an anthropologist, so he was more interested in kinship and living arrangements than in the traits and values that are commonly studied by psychologists. In psychology, similar methods have been described as 'participant observation', referring to the systematic observation of behaviour while living among the group of interest, but keeping a professional distance from the 'objects' of observation. Sometimes, the moral agonies can be too much for anthropologists to bear and they will leave their placement, as in the case of Laura Bohannan (1964), who was confronted with a smallpox epidemic during a year among the Tiv, a primitive bush tribe in West Africa.

Much ethnographic work has used informed insiders as a key source of interpretation and validation of observations. These are typically individuals who work more closely with the ethnographer and help with making sense of the tremendous amount of information that has been gathered. This use of informed insiders in the description of culture has been adopted recently by psychologists studying **intersubjective culture**, which we discuss in more detail in Chapter 8.

Intersubjective Culture refers to shared perceptions of the psychological characteristics that are widespread within and characteristic of a culture.

However, ethnographic methods have their problems. The famous study by Margaret Mead on the sexuality of young Samoan women (Mead, 1928) is widely cited as an example. Mead was said to have imposed her own feminist beliefs on her representation of Samoan society

and sexual norms in ways that could not be sustained by the observed data. Other problems include the status of the ethnographer in the group being studied: can observers truly immerse themselves in a native culture without altering others' behaviour by their presence? Consider Turnbull's decision to feed Adupa. There is also an important distinction between observation and interpretation: to what extent is it possible for an observer ever to realise *a cultural member's* vision of *their* world, as Malinowski proposed?

Other forms of observation may not require participation or informed insiders. Levine and Norenzayan (1999) measured how fast people walked a distance of 60 feet in two downtown areas of major cities in 31 nations, how long it took a clerk to process a request for stamps on a written note at a post office, and how accurate clocks were in public areas. These indicators of 'pace-of-life' were found to be correlated with individualism, wealth and temperature. People in individualistic societies, in rich countries, and in cold climates walk and process economic transfers faster and clocks are more accurate. Since the outcomes of interest in these studies were objective measures of behaviour, they require no interpretation by a possibly biased observer. However, they can tell us little about the lived experience of cultural members.

INTERVIEWS AND FOCUS GROUPS

Interviews and **focus group discussions** are two other widely used methods in cross-cultural psychology that also require interaction between the researcher and cultural informants. Interviews are typically conducted on a one-to-one basis. They can be conducted in a more or less structured manner, ranging from casual interactions between an interviewer and an interviewee to highly structured interviews in which the questions and their order are predetermined. A looser approach allows for a more open exchange of perspectives between the interviewer and interviewee and the possibility of following up any leads provided by the interviewee. The downside of using unstructured interviews is that it can be difficult to compare information across interviews, raising issues of validity and reliability. Obviously, highly structured interviews provide more comparability, but they may be too rigid to explore fully the experiences and viewpoints of the interviewee on a specific topic. Interviews and group discussions are the prime research methods used by indigenous psychologists.

Interviews are used in qualitative research to capture the meaning and significance of particular themes, events, or experiences in the life of individuals by asking each respondent a series of relevant questions. **Focus Groups** are a form of qualitative research in which a group of individuals is asked to discuss their opinions, beliefs, attitudes or perceptions judgments about a particular object, construct, or topic of interest with each other and the researcher.

Structured interviews are more common in cross-cultural psychology. For example, Kärtner, Keller et al. (2007) used a highly standardised form of interview. They presented standardised photographs of child-mother interactions to mothers with 2-month old infants in Berlin and Los Angeles (representative of urban individualistic families), Delhi and various cities in Cameroon (representative of urban and most likely collectivistic families), as well as to

rural Nso mothers in Cameroon (representative of rural and collectivistic families), and asked them to pick and describe their favourite pictures of those presented. As predicted by the authors, mothers from urban (individualist or collectivist) locations referred more to themselves when describing the pictures, whereas mothers from collectivist (urban or rural) locations referred more to other people, to the social context within which the interactions were taking place, and to authorities. (The theory underlying this study is discussed in Chapter 8.)

Focus group research typically involves a researcher interviewing a group of individuals or leading a discussion on a particular topic. As with one-to-one interviews, focus groups can be conducted in a more or a less structured manner; however, because the group members are interacting with each other, the researcher typically has less control over a group discussion than they would over an interview. One advantage of focus group interviews or discussions is that they can reveal areas of contention or multiple interpretation emerging from the interactions among members of the same cultural group or between members of different cultural groups. A potential downside here is that the personality characteristics of the participants (e.g., dominating or shy individuals) may change the dynamics of the interview, and subtle or unrecognised status differences may strongly influence the outcome of discussions. For example, in many traditional societies there are norms and rules about who is allowed to talk first, as well as who may contradict others or provide different information. If there is an individual with higher status participating in a focus group, other members may not speak up and may concur with the views aired by the higher status individual. Nonetheless, focus group discussions are often used in projects focusing on practical issues such as ethnic diversity, discrimination, or mental health among minority employees.

The choice between structured or less structured methods is often linked to researchers' assumptions about reality and about the nature of knowledge. Some researchers prefer structured methods because they assume that there is a reality that can be studied relatively objectively. Where data collection is highly structured, analysis is more likely to involve quantifying the occurrence of particular themes or contents, which can then be subjected to statistical analyses. In contrast, unstructured approaches are typically associated with a greater emphasis on subjectivity and interpretation. Researchers who use these approaches are more likely to conduct purely qualitative analyses, aiming to identify important themes or novel insights through their engagement with what the participants have said, but without seeking to make statistical generalisations beyond the specific context of the study. Research Debate Box 4.1 surveys some of the differing assumptions that researchers can make about reality and about the nature of knowledge.

CULTURAL PRODUCTS AND MEDIA CONTENT

Other options for qualitative researchers are to examine published texts (archival or historical analyses) or cultural products. Such approaches have had a relatively long history in cross-cultural psychology. McClelland (1961) studied children's books from various societies and coded the extent to which a need for achievement was evident in these texts. In Chapter 2, we also discussed a qualitative study of cultural products by Morling and Lamoreaux (2008): these authors

Ontology is the philosophical discussion of being or reality. It is concerned with questions about what entities exist, how we know that an entity exists and what is its meaning, and how any entities may be grouped or related to each other.

Epistemology is a branch of philosophy that discusses the nature, limits and scope of knowledge. It addresses questions such as 'What is knowledge?', 'What is a truth?' or 'How do we acquire knowledge?'.

Positivism is a philosophy of science that is based on empirical observation and verification of data that can be derived from sensory experiences. It is assumed that general laws about the physical and social world can be formulated and that the researcher proceeds in an objective manner to draw conclusions from the collected data.

Postmodernism is a broad philosophical movement that rejects scientific or objective efforts to describe (in our context: psychological) processes. Reality is not independent of human understanding, but is socially constructed by humans. There is no absolute truth, and the way people perceive the world is subjective and shaped by language and power relations.

Post-positivism shares with positivism the meta-theoretical assumption that reality exists, but does not assume that reality can ever be known perfectly. Instead, it is recognised that the process of scientific research, like everyday understanding, is biased by the theoretical background, knowledge and values of the observer or researcher. Karl Popper, an influential proponent of this approach, argued that theories can only be falsified (that is, rejected based on empirical data), and never verified. One important implication is that theories need to be amenable to falsification for them to be assessed scientifically.

counted frequencies of collectivist themes in cultural products, and then related the observed differences to scores on the cultural dimensions identified by Hofstede (2001) and others.

Bardi, Calogero, and Mullen (2009) studied US newspaper content from 1900 to 2000 as a test of the validity of Schwartz's (1992) theory of values (which we discussed in Chapter 2). They developed lists of words that were thought to reflect the ten basic values in Schwartz's theory. They then examined whether newspaper articles across this time period mentioned these values and whether the frequency of mentioned values was correlated with observed behaviours over the same period. They found that values as indexed by newspaper content did indeed relate significantly to behavioural indicators, such as military participation (power), alcohol consumption (hedonism), number of movies released (stimulation), unwed births (conformity) and numbers of police and guards employed (conformity). Thus, they successfully used archival records to examine cultural change in one national culture over time.

The emergence of the internet and new technologies has created opportunities for conducting qualitative research in much broader ways. For example, Boer and Fischer (2012) placed questions about how individuals use music in their daily lives on various internet sites accessed by young people, sampling from New Zealand, the USA and Germany as examples of individualistic societies, and from Hong Kong, the Philippines, Brazil and Singapore as examples of more collectivistic societies. They were able to identify seven main themes of how young people used music on a daily basis, which differed somewhat between the samples. In individualistic cultural samples, respondents reported greater use of music for remembering good times (especially among Germans), whereas in collectivistic settings respondents reported greater use of music for emotion regulation and socialising with friends and family (especially among Brazilians).

Research Debate Box 4.1 Assumptions about Knowledge and Meaning

Research methods differ in the basis of the knowledge that researchers seek. Up to now, we have described some examples of qualitative research methods that can be used in cross-cultural psychology. These methods cover a diverse array, differing in their views of reality (**ontology**) and the nature of knowledge (**epistemology**).

Traditionally, scientists have typically based their research on **positivism**, which assumes that scientific methods are a valid way to describe empirical phenomena in a systematic, objective and logical fashion. Yet this focus on searching for objective descriptions of reality has been criticised. Amongst others, the French philosopher Michel Foucault and the American historian Thomas Kuhn were some of the influential thinkers who have challenged the positivist view, and their critique is now considered part of a broader paradigm of **postmodernism**. Postmodernism encompasses a diverse set of philosophical approaches that stress the subjectivity of experience and emphasise that our reality is socially constructed and therefore subject to change. 'Culture' has been a popular field of investigation among postmodern scholars: these scholars focus on the role of language and power relations in constructing and legitimising particular ideas, categories, or ways of being, rather than the broader patterns of similarity and difference that are more focal within cross-cultural psychology. Central to the postmodern view of reality as socially constructed is that concepts and ideas that are usually taken for granted can be reinterpreted as cultural inventions, and thus they are open to being questioned or 'de-constructed'.

The emergence of postmodern approaches has created many intellectual problems for ethnographers and anthropologists. Some core assumptions of positivist approaches to culture that have been questioned (see Greenfield, 2000, for a summary and response) include:

- that culture can be objectively described by an outsider without the intrusion of the observer's subjective stance on what s/he feels, believes or values;
- that cultures are homogeneous systems;
- that it is possible to describe culture in factual terms without subjective interpretation by a researcher;
- that knowledge is derived from a world that exists independent of the person (i.e., the researcher) knowing it.

Returning to our initial example, a postmodern view suggests that Turnbull's description of the Ik culture was undoubtedly shaped by his own values and beliefs, reflecting his own cultural origins in mid-twentieth century London; that his conclusions about the cultural group as a whole would have been narrowly based on his interactions with selected individuals in the various villages; and that his presence would have changed the actions of the individuals he was observing.

Partly in response to these critiques, philosophers of science have developed **post-positivist** epistemological perspectives (see, for example, Hwang, 2006). Like postmodernism, post-positivism

(Continued)

(Continued)

is not a unitary philosophy. Typically, post-positivists accept that knowledge is contingent on processes of social construction, and thus it is inevitably incomplete and imperfect: however, they do not deny the existence of an external reality, and they tend to view the scientific method as pragmatically useful for describing both objective and constructed realities. In particular, scientific method provides a means of falsification (rejecting incorrect beliefs about reality), and treats human knowledge as not unchallengeable or absolutely true, but open to modification based on further investigation.

Within psychology, postmodernism has been especially associated with a critique of quantitative methodologies. Postmodernist researchers argue that meanings are too fluid and subjective to be quantified or generalised, and thus they tend to adopt relatively unstructured research methods and qualitative analyses involving intensive interpretation. However, many of the quantitative methods we review recognise the impossibility of knowing another person's subjective experience. Grounded in post-positivist assumptions, these methods are based on understanding psychological constructs as 'latent variables' (constructs that cannot be observed directly, but that may be inferred from patterns of observable behaviour, for example responses to multiple items on a questionnaire).

Moreover, from a post-positivist perspective, ideas arising from postmodernism about the culturally contingent nature of beliefs, values, and other constructs can be viewed as theoretical propositions that are amenable to scientific testing. Cross-cultural psychology is ideally suited to test these ideas. For instance, in a study discussed in Chapter 12, Pehrson, Vignoles, and Brown (2009) drew inspiration from postmodern theories of identity in order to derive and test hypotheses about the cultural contingency of the relationship between national identity and prejudice against immigrants.

We strongly encourage research that uses mixed methods, and we would encourage the reader to pay attention to research from different epistemological and methodological perspectives. However, researchers interested in culture need to be aware of these broader philosophical debates and will have to take a position on these matters.

Golder and Macy (2011) examined Twitter posts from around the world and explored whether the way in which people express their mood varies systematically with diurnal and seasonal patterns. They found that, independent of nation of origin, people post more positive tweets in the morning with their mood declining over the course of the day, a finding which fits with biological patterns of sleep and circadian rhythms. People are generally happier at weekends, but are also affected by the relative day length over the course of the year.

One major advantage of these methods is that they tap data that individuals produce either for public consumption (e.g., books, internet web sites and advertisements) or in order to communicate with the outside world (e.g., Twitter or Facebook postings), and therefore they represent culture in the making. Of course, these methods also require carefully prepared coding schemes, highly trained coders and careful interpretation. Golder and Macy (2011) restricted their analysis to tweets in English, and sampling was based on the location from

which tweeters (people who post messages on Twitter) had sent their messages. This sampling strategy could have obscured true cultural differences that would have been of interest to social psychologists and, as we shall see, the use of English may also have led to a convergence of responses.

When examining the use of such cultural products and media content, we need to pause and ask whether the use of English as a universal language of international communication changes the results of studies. Harzing (2005) investigated this problem by sampling students in 24 countries. In each of these, she administered questionnaires to half of the students in their local language whereas the other half received an English language version. Overall, she found that responses in English showed fewer cultural differences, supporting claims that individuals subconsciously accommodate their responses to the stereotypical cultural norms of the language to which they are responding. This effect has been confirmed subsequently (e.g., S.X. Chen & Bond, 2010, discussed in Chapter 5), and used in experimental priming studies, which we discuss below.

Harzing also differentiated between different types of questions. She found the largest effects of language on questions about cultural values and norms, and smaller but still significant effects on more culturally neutral questions, such as reasons for selecting course electives. Consequently, research that uses non-native languages is likely to underestimate cultural differences on more culturally sensitive topics.

Guideline 2: Researchers should study psychological processes in the native language of participants, as using English or other business or trading languages is likely to lead to cultural accommodation of responses to the perceived cultural stereotypes of the language being used, thereby underestimating cultural differences.

PSYCHOMETRIC TESTS

Ethnographic and qualitative approaches have been popular methods for observers of cultural differences for centuries. Researchers looking for more objective ways of describing individuals and groups often rely on psychometric data, in other words the quantitative measurement of knowledge, abilities, attitudes, personality traits, and other such constructs. Rudimentary forms of psychometric testing were developed in ancient China about 4,000 years ago and formalised during the Han dynasty (Gregory, 1996). The most widely used current method for collecting cross-cultural psychological data is by way of paper-and-pencil surveys or questionnaires, as well as online surveys.

These methods are similar to interviews in that researchers seek a response from research participants on questions of interest about their inner feelings, states, personality, beliefs, attitudes, goals, values, or other psychological variables, which usually require some form of introspection. One major difference is that participants are asked to respond using fixed formats, typically with a **Likert-type** response scale requiring respondents to grade their responses along a scale of relative strength of endorsement (e.g., from strongly agree through neutral to strongly disagree). The studies discussed in Chapter 2 all used some form of such questionnaires, as did many other studies discussed in later chapters.

A **Likert Scale** is a rating scale used in questionnaires, named after an American psychologist, Rensis Likert. Respondents are asked to specify their level of agreement or disagreement with a specific statement on a symmetric agree-disagree rating scale (typically a 5- or 7-point graded scale). The response is taken as indicating the intensity of their reaction towards the stated topic.

Reliability refers to the consistency of a measure, both in terms of its component parts and its repeated use over time. A measure has high reliability if it produces consistent results under consistent conditions.

Validity in the statistical sense refers to whether a measure is fit for its intended purpose. Several subtypes can be distinguished. Predictive validity is concerned with whether the measure can predict some other construct of interest (sometimes called utility). Construct validity is the extent to which a measure accurately reflects the variability and relative position of test takers on the underlying construct that the measure is designed to measure.

Objectivity refers to the absence of bias in the measurement process. The measurement process should yield equivalent results independent of the researcher or the instrument used.

Psychometric researchers aim to produce valid, reliable, and objective information about underlying constructs that are not directly observable (e.g., a person's subjective beliefs, or their level of intelligence), by analysing patterns of observable behaviour across a series of tasks (e.g., responses to a series of items on a questionnaire). In single-culture studies, inferences are drawn about the **reliability** of a measure based on careful analyses of the patterns of correlation among individuals' responses to the separate items and/or the similarity of responses on multiple testing occasions. Inferences about **validity** require interpreting how well the contents of the measure match the theoretical construct that the researcher is aiming to measure: this interpretation may be aided by testing correlations of the measure with measures of other constructs that are expected to be related to it (convergent or predictive validity), or that are expected to be unrelated (discriminant validity).

In cross-cultural research, this becomes more complex. A measure that is reliable and valid in one cultural context when transposed to a new context may not yield comparable scores. Moreover, even when the measures are comparable, differences in the testing environment may affect the **objectivity** of measurements. We now consider some of the sources of bias in cross-cultural research, as well as some of the techniques that are available to researchers seeking to make cross-cultural comparisons that are as valid, reliable, and objective as possible.

Sources of Bias in Cross-Cultural Research

Construct bias

One important issue is the extent to which any psychological construct is understood and conceptualised in the same way across cultural groups. Construct bias refers to the cultural specificity of a psychological construct or process. If there is construct bias, the construct is defined differently in two or more cultural groups. For example, greeting procedures are functionally

universal, but structurally non-equivalent (compare kisses to handshakes to bowing). An example of direct practical importance concerns cross-national comparisons of reading skills, as discussed in Box 4.1.

Box 4.1 When is a Difference a Cultural Difference?

National differences in scores of intelligence tests focus on abilities considered without referring to the context. What happens if we make comparisons that do include context? After one year of schooling, children in Finland, Germany, Italy, Spain and Greece all achieve accuracy levels of greater than 90% on reading of words and non-words constructed by the researcher. Children in Portugal and Denmark achieve around 70% accuracy. Children in the UK achieve an accuracy of just 40%. It takes them three or four years to achieve 90% accuracy.

Why is this very large difference found? Is it because children in the UK often start school between the ages of 4 and 5, while those in most other European countries start later? Is it because teachers in the UK are less well trained in the specific skills required to teach reading? Are UK teachers less motivated? The principal explanation for the difference is that English is a much more difficult language to read. It is relatively easy for a Spanish or Greek child to learn to read, because their languages are phonically consistent. In other words, a given combination of characters always has the same sound. In English, vowels are frequently pronounced in different ways depending upon the consonants with which they are paired. Consider for instance the sound of 'a' in 'cap', 'call' and 'car', or of 'o' in 'go' and 'do'.

Comparisons of rates of learning to read differ from comparisons of intelligence scores, because languages differ from one another, whereas intelligence testers attempt to make their tests 'culture-fair'. Should differences in learning to read be considered a cultural difference, or should differences only be attributed to culture when measures are used that are equivalent across cultures? Is language a part of culture?

Sources: Goswami, Porpodas & Wheelwright (1997); Seymour, Aro & Erskine (2003).

A related threat is **domain under-representation**, where a test or questionnaire may miss out important aspects of a construct in a specific cultural setting. The example of the wider definition of intelligence in African communities is one such instance. Qualitative studies where respondents are asked what attributes they associate with intelligence have repeatedly found that intelligence is more broadly defined in collectivist cultural groups (Sternberg, 2007). Intelligence is

Domain Under-Representation is present in a measure if an aspect of the domain that is important to the function of a theoretical variable is missing from that measuring instrument in at least one of the sampled groups.

seen as including social competence in Taiwan, Japan and China. Studies in sub-Saharan Africa have frequently shown that the skills of maintaining harmonious and stable relations within and between groups are seen as key aspects of intelligence, in addition to the cognitive alacrity that is typically understood as intelligence in Western societies. These differences indicate the need for caution in cross-cultural comparisons of intelligence (see our previous discussion in Chapter 3). Box 4.2 provides a telling example of the problems of testing across cultures.

Box 4.2 The Challenges of Conducting Research Across Cultures

A researcher administered an intelligence test to an African child. However, the child sat there mute and did not respond to any of the questions asked by the tester. One of the tasks was to recount the story read out by the tester. As with all the other tasks, the child remained silent and avoided eye contact. Disappointed, the tester finally dismissed the child, convinced that the child had some serious developmental delays in her social and cognitive development. The child was picked up by her caregiver and walked a few hours back to her ancestral village. A few days later another child from that village came for testing. This child was much more forthcoming with answers and was very energetic and talkative. When the tester started reading out the story as for the first child, she interrupted the experimenter and said that she had already heard this story from her friend and started recalling the correct story with all the necessary details. The tester was baffled. The first child had not only remembered the story on the way back to her village, she had also correctly recounted it to other children, including this one now sitting in front of her who memorised it and retold it accurately. Therefore, the first child probably had normal or even outstanding intellectual capacities, but had interpreted the testing situation differently. In her culture it was socially intelligent not to talk back to the tester as an adult, and she had aimed to present herself as an intelligent and wise child that did not talk back to an older person, even if tempted repeatedly.

Source: Adapted from Harkness & Super (1977).

The dilemmas of construct bias can be considered in light of the useful distinction between etic and emic approaches, proposed by Pike (1967) and popularised by Berry (1969). In attempting to learn a Mexican Indian language, Pike found that the use of differing pitches and tones influenced the meaning of specific sounds in that language. In terms of the concepts used in linguistics, phonetic production affects the meaning of specific phonemes. Drawing on this distinction, Berry contrasted two approaches to cross-cultural study. Firstly, one could start from the assumption that there are universals and proceed in that manner until evidence is found for differences. He termed this the 'etic' approach, paralleling the universalist assumptions made in phonetics, the study of sounds. Alternatively, one can start by studying intensively the distinctive attributes of one specific cultural group. He termed this the 'emic' approach, because it focuses

on local meanings, and draws most readily on information provided by persons within that cultural group. This orientation parallels linguists' focus on the phonemic attributes of a specific language.

The debate between emic versus etic methods somewhat resembles the debate between positivism and post-modernism that was discussed in Research Debate Box 4.1. Etic research is strongly associated with positivist and post-positivist epistemologies. Emic research focuses on culture-specific attributes that can only be understood within the local context. The approach used by most indigenous researchers falls under this heading, as summarised in Box 4.3.

Box 4.3 How to do Indigenous Psychology

In Chapter 1, we discussed indigenous psychology as one area of cross-cultural research. Indigenous researchers often prefer to use the less standardised methods that allow for more mutual and egalitarian exchanges between researcher and interviewee (Pe-Pua, 2006). Enriquez (1993) distinguished between two main types of indigenous research: 'indigenisation from within' and 'indigenisation from without'. The former refers to locals or insiders developing a psychology within and for their own culture. The latter describes attempts by foreigners to understand and describe a culture in its own terms, moving from an imposed-etic understanding to an emic understanding (in Berry's terminology). Here are some guidelines that indigenous researchers have used for such endeavours.

Do tolerate ambiguous or vague states of understanding and suspend decisions as long as possible in dealing with theoretical, methodological and empirical problems, until something indigenous emerges in your mind during the research process.

Do be a typical native in the cultural sense when functioning as a researcher.

Do take the studied psychological or behavioural phenomenon and its sociocultural context into consideration.

Do give priority to the study of culturally unique phenomena.

Do base your research on the intellectual tradition of your own culture rather than on that of a Western culture.

Don't neglect Western psychologists' important experiences in developing their own indigenous psychologies, which may be usefully transferred to the development of non-Western indigenous psychologies.

Source: Excerpted from Yang (2000) (emphases added).

Emic studies are often equated with postmodernist concepts, but in fact they can also be conducted within a post-positivist framework. For example, if you develop a testable hypothesis about psychological processes within a specific culture (for instance, 'there will be

relationship between the importance placed on children and the sharing of food'), then you are conducting an emic study within a positivist paradigm. Ferreira, Fischer, Porto, Pilati, and Milfont (2012) conducted such a study. They developed measures of *jeitinho brasileiro,* an indigenous social influence strategy in Brazil (see Chapter 10) and then examined how individual preferences and normative perceptions of *jeitinho brasiliero* related to other psychological variables.

Berry (1969) suggested that cross-cultural research often starts as 'imposed-etic', that is to say, it is based on applying Western concepts and measures in non-Western contexts. Imposed-etic research relies on the assumption that the concepts and measures will have the same meaning in new contexts. Rokeach's (1973) study of social values using a survey developed in the USA is a good example of emic US research that was subsequently 'imposed' in nine other nations (Ng et al., 1982, discussed in Chapter 14).

As a research field becomes more fully developed, an accumulation of emic studies can contribute to the development of improved 'derived-etic' measures, that have equal validity in a broad range of cultural contexts. Rokeach's work strongly influenced Schwartz's (1992) subsequent cross-cultural studies of values. His survey items were drawn in part from Rokeach, but he also consulted Muslim and Druze scholars and included values from previous instruments developed in China, Israel and Zimbabwe. Moreover, he allowed for the possibility that he might be missing further values, by asking his collaborators in each country to add culture-specific values of local relevance. However, the local items proved compatible with the universally identified dimensions, so his individual-level value dimensions can be considered the best current example of a derived-etic measure encompassing a broad range of cultural groups.

Instrument bias

Even if an underlying construct can be defined equivalently across cultures, it may not be measurable in the same way. Serpell's (1979) study of British and Zambian children provides a good example of instrument bias in the area of cognitive abilities. British students were much more familiar with using pencils (or pens) and paper in their daily lives, and when standard pencil-and-paper tests were used the British outperformed Zambian children. In contrast, if iron wire figures were used, Zambian children showed superior test performances, since making toys from iron wire was a common pastime in Zambia and children were familiar with performing these tasks. Importantly, the two groups did not differ on a task involving clay figures that was unfamiliar to both cultural groups.

Let us reconsider some of the studies discussed in Chapter 3, for example the study by Henrich et al. (2010) who sampled participants from 15 traditional cultures around the world and asked them to play a number of economic games. Imagine you are a Hadza forager who is roaming the Tanzanian savannah and is used to collecting and hunting for the food that is necessary for survival. You do not use money on a regular basis, although you are probably aware of its existence. Now you are interacting with some foreigners who play a game where you have to imagine that a person that you do not know is giving you a large amount of this money and you have to divide it between yourself and an anonymous person that you have never seen and indeed never will. In their study, Henrich et al. reported that the Hadza were among the

least likely to offer any money to the anonymous player. They explained differences across communities by showing that integration into a market economy and believing in a monotheistic god are necessary for the emergence of fairness (operationalised as sharing in the so-called Dictator Game). This is an imposed-etic interpretation and we could equally argue that the results test instead for a familiarity with Western concepts such as money and abstract, non-personal notions of fairness that are not relevant in a Hadza community.

Translation-Back Translation is the classic translation procedure. An instrument is first translated from the source language into the target language and then independently translated back into the source language. The original and the back-translated versions are then compared and changes are made so as to improve accuracy.

Decentering involves replacing culturally specific expressions in the initial version of an instrument with alternative wordings that are more translatable but still preserve the underlying meaning.

It is preferable to use locally meaningful and relevant objects for exchange games like this one. A good example is a study by Apicella, Marlowe et al. (2012) which also involved Hadza participants. They studied exchange networks in Hadza camps by giving participants three sticks of honey (a prized food source among the Hadza) and then asked them to distribute these three honey sticks in any way they liked among people in their camp. Of course, this was a study within a single cultural group. In a cross-cultural study, locally equivalent objects would need to be found so that the results could be compared across groups.

The creation of measures that are understood equally well and in similar ways across different parts of the world is not simply a matter of using items that refer to issues or tasks that are familiar to respondents. There is also a need to ensure that translations from one language to another are done in a manner that yields items with equivalent meaning (Hambleton & Zenisky, 2011). The most widely accepted procedure for achieving linguistic equivalence is **back-translation** (Brislin, Lonner, & Thorndike, 1973): a translation is first made from the language in which the test was originally developed into the language of the society in which it is to be used. A second bilingual person is then asked to translate the items back into the original language, usually English, without having seen the original version. A comparison of the retranslated version and the original by a native speaker-writer of the first language can then be used to detect problematic translations and to create an improved version through discussion between the two translators.

Back-translation is cost-effective and fast, but can result in stilted and awkward sentences in the target language. It is therefore not the optimal solution. A 'committee' approach, where a group of bilinguals get together and discuss the meaning and appropriateness of each item, may overcome this problem. It is possible to combine back-translation and committee approaches to capitalise on the strengths of each method. However, such methods are very labour and time intensive (see Harkness, 2003, for more details).

Difficulties often focus on the relative merits of a literal translation versus linguistic '**decentering**'. A decentered translation does not necessarily use terms that have a precise linguistic equivalence, but may draw on the cultural knowledge of the translators to use phrases that have

an equivalent meaning in the two languages. For instance, while English speakers discussing some misfortune might seek hope by claiming that 'every cloud has a silver lining', speakers of Mandarin Chinese would claim that 'every cloud has a pink edge'. A further step in decentering would be to drop the specific descriptors in the original version in favour of a similar, more general saying, like 'something good comes from any misfortune'. A good example of decentering was the development of the social axioms study discussed in Chapter 2; Leung, Bond et al. (2002) gathered proverbs, maxims and adages in Hong Kong and Venezuela, but then had to rewrite them to capture the underlying social beliefs in more general terms that would be understandable outside their cultures of origin.

Key Researcher Box 4.1 Kwok Leung

Figure 4.1 Kwok Leung

Kwok Leung (b. 1958) grew up in Hong Kong. After discovering cross-cultural psychology while taking an undergraduate course at the Chinese University of Hong Kong with Michael Bond, he obtained his PhD in social and organisational psychology from the University of Illinois, Urbana-Champaign, supervised by Alan Lind and inspired by Harry Triandis. He is currently chair professor of management at the City University of Hong Kong. He has collaborated with Michael Bond in pioneering cross-cultural studies of beliefs (social axioms) and has conducted many studies of justice. He has also co-authored key publications on cross-cultural research methods. Overall, he has published several books and more than 100 academic articles.

Administration bias

Administration bias refers to any aspect in the administration of a test or survey that affects results. This can include differences in physical conditions and social environments (e.g., the noise or temperature of the test setting, or the presence of other people in the test situation), instructions (e.g., the different experience of test administrators in different contexts or the use of vague language in instructions), the status of the test administrator (e.g., their ethnicity, profession, status, religion, or local versus foreign status) and communication problems (e.g., differences in the language used, the use of interpreters or the culture-specific interpretation of instructions). The goal in classic psychometrics is to standardise the test situation as much as possible, so as to rule out any variation in the results that could be interpreted as arising from differences in the administration of the scientific procedure.

However, from a cultural perspective this standardisation may introduce error variance if the required procedures result in a situation that is not compatible with local customs and cultural standards. Practitioners of ethnomethodologies such as *pagtatanong-tanong* (roughly translated

as casual asking around) in the Philippines (Pe-Pua, 2006) have criticised Western psychologists for striving to standardise social situations and limit the interactions between participants and researchers. One of the key aspects of *pagtatanong-tanong* as an alternative to Western psychological procedures has been the concept of casual and non-directed conversations that are driven by the respondent rather than the researcher. It is through these conversations that the researcher can obtain valuable information once a certain level of trust has been established and the researcher is no longer treated as an outsider.

Nevertheless, these unstructured approaches have been heavily criticised as yielding unreliable and non-replicable results in cross-cultural psychology (Van de Vijver & Leung, 1997) and similar criticisms have also been raised within the Philippines. Pe-Pua argues that these problems can be overcome through a systematic documentation of findings, the replication of studies, and giving attention to ethical guidelines. Indigenous methodologists stress that the relationship takes priority and that interview procedures cannot be guided by the demands for standardisation required by positivist methodologies. According to this approach, any strict standardisation would feel alienating to participants in non-Western contexts.

Key Researcher Box 4.2 Fons van de Vijver

Figure 4.2 Fons van de Vijver

Fons van de Vijver (b.1952) grew up in the Netherlands, and completed his PhD at the University of Tilburg. He is currently a professor at the same university and holds additional posts at North-West University, South Africa, and the University of Queensland, Australia. He has published 16 books and more than 350 papers and book chapters, mainly in the domain of cross-cultural psychology. His main research topics involve measurement bias and equivalence, psychological acculturation and multiculturalism, cognitive similarities and differences, survey response styles, and translations and adaptations of psychometric tests. He has co-authored key publications on cross-cultural research methods.

In the context of developmental psychology, Abubakar (2008) has suggested that tests that require strict standardisation can be administered last to overcome some of these limitations when testing children in African contexts. For example, the test administrator may start off by playing football with the child and her siblings, which allows for observations of balance, control and other aspects of motor function that are important for assessment. Once the child has developed some trust and feels comfortable with the test administrator, more standardised instructions to test cognitive abilities can be used. This graduated

approach circumvents problems associated with strict standardisation that may alienate respondents.

Sampling bias

Finally, sampling bias refers to the characteristics of the sample that is used for comparisons. Reconsider the discussion of the Ik at the beginning of this chapter. What sample should be used against which to compare the values and moral inclinations of the Ik? The Ik groups studied by Turnbull varied in so many different ways from most other samples studied by psychologists that it would be hard to find a comparable group. Cross-cultural studies frequently compare student samples, but students are drawn from elite groups in some nations much more than in others. Enrolment in tertiary education varies between 73% in the USA and 7% in China. Even in the 'Western' nations of Europe it varies between 70% in Finland and 30% in the Czech Republic (www.nationmaster.com).

In many of the studies reviewed in this book, psychology students were the research participants. You may ask the question to what extent do psychology students differ across societies in ways that are additional to their cultural background? One important difference can be the status of psychology in a given society, and the educational requirements for studying psychology there. If you are a student residing in Germany, you probably went through a central allocation system before being admitted to a university that takes into account your grades from high school. In Germany, there is a high demand to obtain a degree in psychology, so entry to university is competitive and few students will get their choice of university. In the UK and the USA, psychology is similarly attractive to students, but few constraints exist on choosing to study psychology rather than another major. Psychology students therefore differ between these three nations on important cognitive and motivational variables. In many non-Western societies, psychology has much less prestige and there is no competition for places. Another difference is the socio-economic status of students in each society. These factors are likely to influence motivations, attitudes and beliefs in addition to any cultural variable that may influence survey responses.

As a general rule, the more culturally and economically diverse the samples, the more likely it is that one or more biases will affect the findings. This raises the probability of finding differences that cannot be explained by a single or limited set of variables. This awareness leads to a so-called '**interpretation paradox**' (Van de Vijver & Leung, 2000). It is relatively easy to find differences between samples that differ along many different social, cultural and economic dimensions, but it is then more difficult to pinpoint why these differences exist and what factor or factors can explain them.

The **Interpretation Paradox** refers to the problem that psychological differences between samples that vary along many different social, cultural and economic dimensions are easy to find, but it is then difficult to explain why these differences exist and what variables may cause them.

Guideline 3: Samples should be matched as closely as possible in order to rule out alternative explanations for observed differences in the outcomes being studied.

The Quest for Cross-Cultural Equivalence

As we have just seen, there are many challenges in making valid comparisons between groups. How confidently, then, can we use psychometric measures of values, traits, or abilities to test reliably (consistently) and validly (truly) whether two different populations are the same or different on the value, belief, trait or ability in question? Cross-cultural methodologists have developed guidelines for establishing four increasing levels of equivalence in measures of psychological processes and constructs (Van de Vijver & Leung, 1997; Fontaine, 2005; Matsumoto & Van de Vijver, 2011). The higher the level of equivalence, the more confidently we can claim to be measuring the same underlying construct across different cultural samples. Here, we use terminology adopted by Fontaine (2005). Van de Vijver and Leung (1997) put functional and structural equivalence together as construct equivalence.

Functional equivalence

Functional equivalence refers to the situation where the same theoretical variable accounts for the same measurement outcomes across cultural groups.

The most general and broadest level of measurement equivalence is **functional equivalence**, that is whether the same underlying psychological construct can be said to exist across different cultural contexts. In operational terms, can the test behaviour (e.g., the behavioural response of selecting a response category to a Likert-type item) be seen as an expression of the same theoretical variable? This issue is most easily described and discussed in relation to cognitive phenomena such as intelligence. As we have seen, intelligence is consensually defined more broadly or more narrowly in different cultures. The ability to solve abstract cognitive problems, typically emphasised in Western definitions of intelligence, is only one aspect of the much broader conceptualisation of intelligence that prevails among various African groups.

Functional equivalence is the most abstract and difficult level of equivalence to establish, as it requires an in-depth understanding of each cultural context and extensive qualitative and conceptual work. It is often assumed rather than tested, because it depends additionally on philosophical and theoretical considerations, and no statistical tests are available for making such judgments. Qualitative and ethnographic research is necessary if one is to make an informed judgment about the nature of a construct in each culture and thus to make claims about construct equivalence.

Structural equivalence

Structural Equivalence refers to the situation where the same measurement instrument is a valid and sufficient indicator of a theoretical variable of interest to the researchers in two or more cultural samples.

If functional equivalence is considered to be tenable, the second issue is whether the same items can be used in different cultural contexts as indicators of the underlying construct. This level of equivalence is called **structural equivalence**.

To illustrate the use of comprehensive procedures for establishing both functional and structural equivalence, we describe here a cross-cultural study of emotion research. Breugelmans and Poortinga (2006) were interested in the hypothesis that all humans have the same physical and biological capacities to experience certain emotions, even though these capacities cannot be expressed through language in settings where the language does not have words for specific emotions. For example, members of the native American group of Raramuri in Mexico use the same word for 'shame' and 'guilt'. Does the absence of a specific word indicate that they cannot differentiate and respond appropriately to different social situations that would elicit either shame or guilt in another culture?

To assess this question, Breugelmans and Poortinga started by soliciting descriptions of typical situations that invoked either shame or guilt from Javanese villagers in Indonesia, chosen because they are comparable in many socio-economic characteristics to the Raramuri. They next found that both Dutch and Javanese students were able to differentiate these two types of scenarios. They then presented these scenarios to Raramuri villagers, asking them how they would feel in each situation. The results showed that they responded to the two kinds of situations in a very similar way to that of the Javanese. This demonstrates that the Raramuri were able to discriminate between culturally relevant scenarios of guilt and shame in very much the same way as other groups that do have separate words for shame and guilt. Hence, the non-availability of a word does not affect the differential emotional experience of shame versus guilt.

This example demonstrates that using culturally relevant scenarios can establish both a functional and structural equivalence of psychological processes, even if the observed behaviour (the lexical encoding of emotions) would indicate otherwise. Thus, the critical way to establish structural equivalence is to find indicators that tap the construct of interest in a culturally meaningful way. Indicators need to be relevant and representative of the construct in each cultural setting, but may vary between groups.

To develop meaningful indicators in each culture, extensive qualitative and ethnographic research is necessary, followed by sound psychometric analysis of the instrument. Unfortunately, many published cross-cultural studies have not followed these requirements and have instead used imposed-etic measures. Often, much work is spent in examining the simpler task of whether survey items load in psychometrically similar ways on the expected construct, but far less effort is dedicated to developing locally relevant measures. A major problem is that important aspects or indicators of the psychological construct are often missed when imposed-etic measures are used. This failure can be expected to lead to domain-underrepresentation.

Achieving structural equivalence suggests that the same items can validly be used to measure individual differences in the same underlying construct within both groups. However, this does not yet mean that one can validly compare correlational patterns or mean scores across groups. Before doing this, further levels of equivalence are required.

Metric equivalence

To compare correlational patterns across groups validly, **metric equivalence** is needed. Metric equivalence means that a difference of one scale point on a measure can be assumed to reflect the same difference in the underlying construct across groups. In the psychometric literature this is often called metric invariance. Because we cannot directly measure the underlying construct, metric invariance is inferred from the patterns of relationships among the items. If factor loadings for all items are found to be equal across groups, this supports the assumption that the items are related to the underlying construct in the same way. As the level of an individual on the **latent variable** (often defined as a latent factor) increases, so too does his or her endorsement of items in the questionnaire or test and, crucially, metric invariance suggests that this relationship between latent variable and observed indicator is the same in all cultural groups studied.

Cross-cultural tests of metric invariance are often used to identify and remove problematic items from a scale. Panel (a) in Figure 4.3 shows an item that is performing acceptably for metric equivalence. In both groups, an increase in the latent variable is associated with the same increase in the observed variable. Compare this result with that in Panel (b), where the increases in group 1 are much larger than those in group 2. For example, the item 'working hard to achieve good grades in school' may be a valid indicator of achievement values in literate societies where achievement in school is a strong indicator of achievement, but it might not be as valid an indicator in a pastoral community where children may go to school but achievement is evaluated against non-school criteria. This differential relationship between the observed indicator and the latent variable is called **non-uniform bias**. The term 'non-uniform' emphasises that changes in the latent variable are not uniformly associated with changes in the observed item.

Metric Equivalence refers to a situation where relative comparisons (e.g., mean patterns or correlations) are possible between two or more cultural groups. This result indicates that items have identical relationships with the latent underlying variable in all cultural groups.

A **Latent Variable** is a hypothetical variable that is not directly observable, but can be inferred from other variables that have been directly observed or measured. The inference of a latent variable is done through mathematical models, such as factor analysis.

Non-Uniform Item Bias is characterised by individuals not showing the same ordering on the measurement instruments as would be expected based on their ordering along the latent variable. Bias is present and the size of this bias for a respondent in a group depends on the position of that individual on the latent variable.

Uniform Item Bias is characterised by individuals in two cultural groups showing the same order on the observed measurement that corresponds to their ordering on the latent variable. However, there are some relative differences between the two groups that are not accounted for by the latent variable. The degree of bias is the same for all individuals.

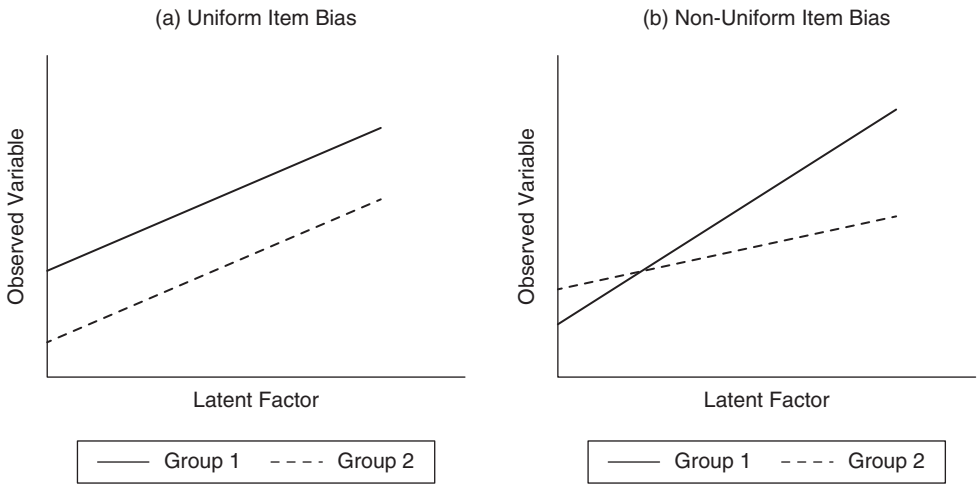

Figure 4.3 Examples of Item Bias

Psychometric tests for metric equivalence are available and discussed in detail elsewhere (Van de Vijver & Leung, 1997; Fischer & Fontaine, 2011). If we find metric equivalence, we can compare patterns of correlations across cultural groups. Differences within each group can be analysed, but absolute mean differences cannot yet be compared across groups. As is seen in Panel (a) in Figure 4.3, the means for group 1 are consistently higher independent of the latent variable. This difference is due to other factors independent of the latent variable of interest. This difference is called **uniform bias**, indicating that there is some difference that is not due to the latent variable (hence the term 'bias'), but this difference is constant (uniform) along the levels of the latent variable. However, the presence of bias means that we cannot make a comparison of scores and attribute them to the latent variable of interest. Only their correlations may be compared across groups.

Full Score or Scalar Equivalence refers to a situation where the scores can be directly compared between two or more cultural groups. It assumes that the measure taps the same base level or intercept in all cultural groups.

Scalar equivalence

The final level of equivalence is called **full score** or **scalar equivalence** (or scalar invariance in the psychometric literature) and allows us to compare group means directly. Full score equivalence requires that the scale is not affected by uniform bias: a score of 5 in one group indicates the same level on the underlying construct as a score of 5 in another group. As with metric equivalence, a full score equivalence cannot be tested directly, because we cannot directly measure the underlying construct. However, we can infer that particular items may suffer

from a uniform bias if they show differential mean levels in relation to the other items across the groups being studied. In statistical terms, researchers can test whether the item intercepts, that is whether the points where the regression line of the latent variable on the observed indicator crosses the y-axis are identical or not. In other words, is the baseline of an instrument comparable?

This level of equivalence is the most difficult to establish, but is necessary if we want to make inferences about the mean levels of the latent psychological trait of interest. Unfortunately, until recently very few researchers have fully tested levels of equivalence and in particular full score or scalar equivalence. This means that many of the studies reported in this book are open to alternative explanations.

One of the most challenging of these explanations, as we mentioned in Chapter 2, is acquiescent responding. If people agree with an item more strongly than they actually believe, we have the pattern seen in Panel (a) in Figure 4.3. In this situation, the mean score reflects both the variable of interest *as well as* the group's average tendency towards acquiescent responding. To rule out such alternative explanations, researchers need to evaluate their data properly and use methods that can rule out this potential confound. This also applies to other aspects of response style such as extreme responding or moderate responding, where these are found to be culturally distinctive.

Guideline 4: Instruments need to be valid and reliable in all cultural groups in order to accurately detect cross-cultural similarities or differences. Researchers should try to ensure that instruments, measures, and manipulations are understood in comparable ways in each location.

EXPERIMENTS

Internal Validity refers to the level of confidence that a researcher can place on the premise that changes in a dependent variable are caused by the experimental manipulation alone and that alternative causal explanations can be ruled out.

Up until now we have been dealing with methods that psychologists share with many other social science disciplines. The basis upon which psychology became distinguished from neighbouring social sciences was through its emphasis upon studying samples of individuals within controlled settings, rather than focusing upon larger groups, organisations or nations. Experimentalists test the specific effects of changes within controlled environments to which individuals have been randomly assigned so as to rule out alternative explanations. The issue of **internal validity** (establishing that changes in the dependent variable can only be attributed to the experimental manipulation) has been a primary concern for psychologists trying to establish our discipline by emulating the physical sciences in their rigor and objectivity.

Obviously, we cannot randomly assign individuals to cultural groups. However, culture is often treated as a quasi-experiment, as people are born and socialised into different cultural

systems. Thus researchers often use the logic of experimental psychology to infer 'effects of culture' by comparing survey or test scores between samples from two different nations. Of course, individuals will also differ on a number of other dimensions that fall outside our definition of what is cultural (e.g., wealth and education) or are indirectly related to cultural variables (religion and political ideologies). This requires the experimenter to select tasks that can rule out alternative interpretations.

As we discussed in Chapter 1, one of the earliest applications of experimental methods to the charting of cross-cultural differences in psychological functioning was conducted by William Rivers (1901), investigating perceptual processes among Torres Straits Islanders. Today, experimental studies typically involve participants from two or a few cultural groups, most often comparing East Asian and US or other Western students. Participants are assigned to different experimental groups on the basis of their cultural background and will then perform one or more brief tasks designed to test explanations of the cultural contrasts that are found. These approaches have been used frequently by cultural psychologists residing in the USA and Canada. For instance, Peng and Nisbett (1999) compared Chinese (most of whom were studying in the United States) and US students in their preference for dialectical proverbs containing apparent contradictions. They presented dialectical and non-dialectical proverbs to students and asked them to rate how much they liked each proverb. They found that Chinese students preferred dialectical proverbs more than non-dialectical proverbs, compared to US students. They then asked their respondents to evaluate brief reports on pairs of scientific studies that appeared contradictory. The Chinese were more willing to conclude that both studies had merit, while the Americans were more disposed to judge the conclusions of one study correct and the other false. Experimental studies of this kind are frequently used to study contrasts in social cognition, and we review a broad range of such research in Chapter 6.

Studies of this type are quasi-experiments, because the independent variable that is hypothesised to represent cultural differences cannot be randomly assigned to respondents. They are simply categorised (e.g., as Asians versus Westerners), often without measuring the cultural dimensions or processes expected to explain the observed differences or paying attention to alternative explanations. Although these studies do appear frequently in mainstream social psychology journals today, they mark a return to what Bond (2009) called Aristotelian research (see Chapter 1). The theorising and methodological rigor is arguably richer than in equivalent studies conducted in the 1970s and 1980s, but unless some proposed psychological mediator is included and measured, we still cannot make any firm interpretations of cultural effects. The cultural background of the participants is the crucial variable that moderates the effect of an experimental manipulation, but it often remains 'packaged'. In other words, it is unclear what specific aspects of the differing cultural backgrounds of the samples are causing these effects.

It may well happen that an experimental manipulation is not as effective in one culture as in the other. Any claims about psychological processes involved therefore hinge on the success of the experimental manipulation, that is, on the researcher's ability to show that the

manipulation was effective in both contexts. The better experiments are therefore the ones that include both an experimental manipulation check and measures of the psychological process variables that help to explain (i.e., unpackage) the cultural differences and rule out alternative cultural explanations.

In short, experimental studies select samples based on underlying cultural dimensions that are thought to influence the dependent variable of interest. Experiments in cross-cultural settings are quasi-experimental because participants cannot be randomly assigned to experimental conditions. They are always open to alternative explanations.

Guideline 5: Experiments are more persuasive in their evidence if a) they include a manipulation check on the effectiveness of the crucial manipulation in all cultural samples, and b) the experimenter specifies the relevant psychological processes and relates them to the dependent variable of interest, to rule out alternative theoretical explanations for the observed cultural difference.

Priming Studies

A different group of experimental studies has focused on eliciting or 'priming' cultural processes rather than relying on country of origin as an independent variable. Priming is thought to temporarily activate procedural or tacit knowledge (cultural mindsets) and mental representations (including beliefs, values, goals or norms) in people's minds. The evoked mindset then serves as an interpretive frame for processing any subsequent information and switches on heuristics and processing strategies that are relevant and effective in the given context. Priming is therefore a short-term equivalent to the long-term 'chronic' activations that are produced by cultural products and cultural agents in one's socialisation environment.

In these studies, individuals typically engage in a series of tasks (Oyserman & Lee, 2007, 2008). In a first set of tasks, individuals are required to engage in a brief activity during which particular psychological concepts, knowledge or motivational goals are activated. They then participate in another task, in which the cues made salient in the previous activity are carried over and now influence the behaviour on the second task. The purpose of the first task is to prime particular responses which will then spill over into the next, apparently unrelated task, in which the actual dependent variable is embedded.

Given the widely reported differences between individualistic and collectivistic cultures, the various primes deployed typically focus on making either the individual or group mindset typical for another cultural sample salient in people's minds. Researchers randomly assign participants to one of two conditions, in which either individualism or collectivism is primed. Alternatively, one of the two primes is compared with a control condition of no priming. This between-subject design is intended to simulate salient cross-cultural differences. The idea is that priming targets the active ingredient of culture and can explain why previously reported differences in behaviour were found across national cultural groups. Researchers have sought to prime culture either by priming national symbols or languages, which we discuss in Chapter 6, or by priming

the content of cultural dimensions such as individualism-collectivism or self-construals, which we discuss in Chapter 7. These studies can provide interesting information about processes that may help us to unpackage cultural differences but, as we discuss more fully in Chapters 6 and 7, their ecological validity and their explanatory reach have sometimes been greatly exaggerated.

Advocates of priming studies imply that the early focus of theorists on defining culture in terms of nation-level constructs can be dispensed with. For these researchers, culture is a mindset in our heads and is open to continuous and variable elicitation by everyday momentary events. In terms of our present discussion of research methods, the key issue raised by the work of these researchers is whether experimental primes are a striking instance of domain under-representation. Do the effects that they elicit fully represent the cultural differences that they seek to explain?

INDIVIDUAL-LEVEL AND NATION-LEVEL EXPLANATION

As we have stressed repeatedly, culture exists between individuals and would not exist if each human being was on some proverbial island, finding food by him- or herself, and only meeting a partner once a year to mate. Instead, humans live as a part of complex social systems, full of norms, taboos, and toys that allow us to fly to the moon, meet a romantic partner at the cinema, and find out the latest gossip on Facebook, Google+, Orkut or whatever other social network is currently in fashion. A continually intriguing question is whether the aggregated or averaged information gained from individual-level psychological measures from national samples can tell us anything about the norms that seem to influence us on a daily basis. Hofstede's (1980) lasting legacy was to take the bold step that led to the popularising of nation-level dimensions of culture. Thinking about nation-level dimensions raises two questions that need more detailed consideration here, namely whether instruments and constructs at the nation level still have the same structure as at the individual level (**isomorphism**) and whether the relationship between the psychological construct of interest and a third variable is the same at the nation level as it is at the individual level (**homology**). These two questions of isomorphism and homology are often confounded and difficult to answer.

Isomorphism is the extent to which a psychological construct, and the instrument that measures the construct, have the same meaning and dimensionality (i.e., 'internal' structure, for instance, factors) at the individual and the nation levels.

Homology is the extent to which a psychological instrument has the same relationship with an external variable ('external' structure) at the individual level and at the nation level (using aggregated individual scores).

In Chapter 2, we discussed the results of Schwartz's work, in which he emphasises the importance of the differences between the individual- and nation-level organisation of values. Despite striking differences in the positioning of some individual items, the overall value

structures he obtained at the two different levels do look remarkably similar. Nevertheless it is important to maintain a theoretical distinction between these two levels as Schwartz (2010) maintains, and as we now explain. Cultural groups such as nations cannot be compared using individual-level scale means, and likewise, individuals within nations cannot be compared using nation-level value scales. Comparing the values endorsed in Germany, Kenya and Argentina, we need to use the means based on nation-level dimensions because we are dealing with representations of national groups rather than individuals. Using scores based on individual-level dimensions would confuse within-country and between-country variability. For instance, since the Schwartz values 'humble' and 'social power' refer to conflicting motivational orientations for individuals, we cannot compare individuals using a combination of uncorrelated items such as these two. These items refer to different and conflicting underlying latent dimensions (tradition values and power values). However, at the nation level, a statistical combination of values such as 'humble' and 'social power' into a single index does make sense because these values are positively related at the nation level and therefore do refer to the same underlying latent construct (hierarchy). Confused measures lead to confusion in results, because means can only be validly interpreted at the appropriate level. This point underlines a subtle, critical, yet largely unrecognised implication of the lack of isomorphism between individual and group levels of value organisation. We can compare individuals within nations using the individual-level dimensions described by Schwartz (1992), but we need to use the country-level dimensions (Schwartz, 1994) if we want to compare samples from different nations with each other.

The problem of homology is equally interesting and complex. For example, Pehrson, Vignoles, and Brown (2010) measured prejudice and identification with one's nation in samples from 31 nations. At the individual level, prejudice was weakly but positively associated with national identification. However, at the nation level these variables were strongly negatively correlated: in other words, nations with higher levels of national identification had lower levels of prejudice, the opposite pattern. For a more technical understanding of this knotty issue, see Box 4.4.

In practical terms, researchers have constructed nation scores by taking one of three options. Firstly, they have simply used the averaged scores of individuals on metrically-equivalent constructs within each nation as an indicator of its culture. We refer to scores of this type as a citizen mean. This option was taken by Leung and Bond (2004) in their social axioms study. Secondly, some researchers have aggregated items from all individuals within each sample to the nation level and then created nation-level versions of variables that had been validated at the individual level. This option was taken by the GLOBE researchers (House et al., 2004). Thirdly, some researchers have aggregated all items from all individuals within each nation and then examined the structure of items across nations. This last option was taken by both Hofstede (1980) and Schwartz (2004). Options 1 and 2 do not give us answers about the similarity of individual- and nation-level structures, because the individual-level structure was used at both levels. Only option 3 can allow us to compare the resulting structure with that found at the individual level.

Interpreting non-equivalent results

What are the implications of non-isomorphism for tests of invariance at the individual level? If the structure of a scale is not isomorphic across levels of analysis, this implies a failure to find full scale equivalence, because the group means across items differ in ways that cannot be accounted for by the individual-level structure of the items (see Figure 4.3). How, then, can we interpret the cultural dimensions?

A strict interpretation of the relationships between equivalence and isomorphism is that dimensions that have different structures at the individual and nation levels indicate artifacts of measurement bias at the individual level. A possible psychometric interpretation is therefore that these dimensions cannot be interpreted (and should not be used), because they reflect measurement bias in at least one or more samples.

A more lenient interpretation is that bias is of some relevance here because it shows that cultures operate differently. Therefore 'bias' can become a variable of interest. With recent advances in both theory and methodology, it is now possible to study this bias and identify the cultural dynamics leading to the differences in structure between samples. For instance, we can explore the tendency to respond acquiescently as a dimension of cultural variation (Smith, 2011a), and how the level of the human development index introduces bias into the structure of Schwartz's measure of universalism values (Davidov et al., 2012). Future research that explores when and how different structures emerge opens up a new field of research and can provide fascinating insights into how the context within which individuals are operating influences their psychological realities (see for example, Fontaine, Poortinga et al., 2008; Fischer, Ferreira et al., 2011; Fischer, Milfont & Gouveia, 2011).

A third viewpoint, which is advocated by more sociologically-minded researchers and by many cross-cultural researchers interested in national dimensions of culture, is that an exploration of nation-level dimensions is valid in its own right. This is because we are dealing with the properties of a social aggregate, which is captured in the average responses of individuals. Any bias found at the individual level is a kind of evidence of the 'existence' of cultural differences. Remember also that the nation-level correlation component is based on mean scores and is thus statistically independent from the individual-level correlation component, and can therefore be validly explored in its own right. What is crucial about this last position is that we are not dealing with analyses that are explicitly psychological any more. As Leung and Bond (2007) put it, 'eco-logic' is not the same as 'psycho-logic'. What is found at the nation level cannot be meaningfully linked directly to individuals. It is an aspect of the context within which people live (Schwartz, 2010). For instance, power distance refers to nations, and not necessarily to individuals' attitudes to power. This discussion has focused on isomorphism, but the same principles apply to questions of homology (that is, the similarity of the relationship between constructs at individual and nation levels).

Multi-level Modeling

Data that have been aggregated to the nation level, averaged within each nation and factor analysed yield dimensions at the nation level. We can then explore correlations with other indicators

Box 4.4 Portraying Non-Isomorphism

Let us step back for a second and consider the basics of what we are doing. Mathematically, the correlation in any sample can be broken down into two independent parts, a within-group (within-nation) and a between-group (between-nation) component. This is similar to any of the discussions of ANOVA that you will have seen in methods and statistics textbooks. These two components further entail information about the variability within and between groups and the correlation within and between groups.

Technically, it is possible that we will encounter different structures and correlations at the individual and nation level, leading to non-isomorphism and non-homology. Figure 4.4 shows this situation graphically. As you can see there, the regression in each group separately is positive. However, because the means between the groups are different, the correlation of the means (that is the nation-level correlation) will be different compared to the correlations within each group separately (Leung & Bond, 1989).

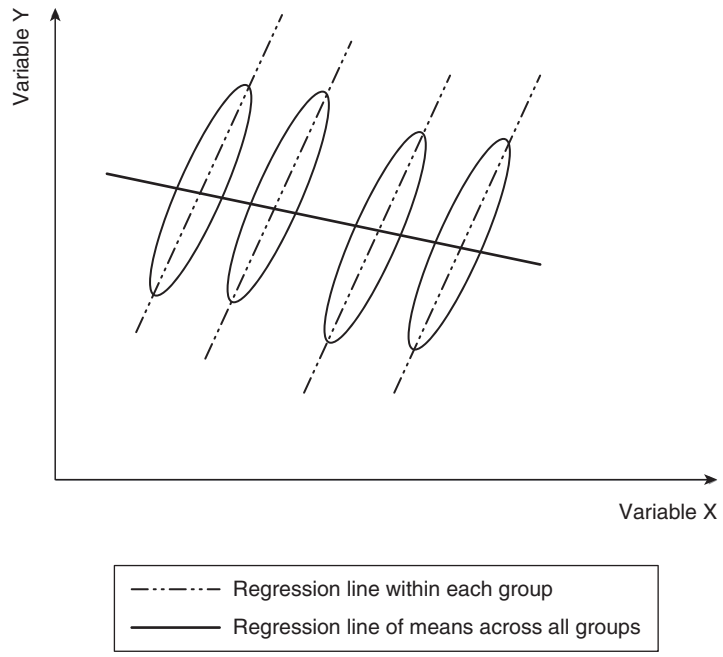

Figure 4.4 Effect of Within- and Between-Group Differences on Correlations

to establish their validity, as Hofstede (1980) did. Tables 2.2 and 2.3 showed some of the correlations that have been found between prominent nation-level dimensions.

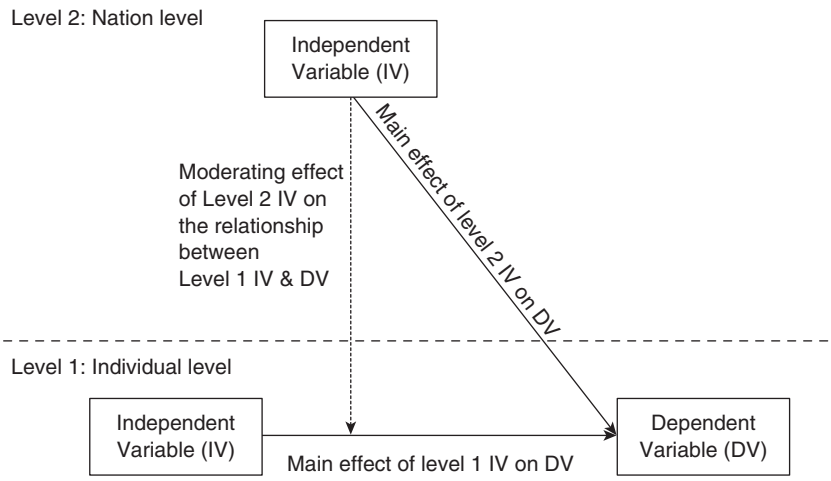

Figure 4.5 Example of a Hierarchical Linear Model with a Moderation Effect of the Level 2 Variable

More interestingly, advances in statistical methods now allow us to test whether these aggregated dimensions have an influence that goes back onto psychological processes at the individual level, while controlling for the corresponding variables at the individual level. We can therefore now truly examine the relative importance of individual versus aggregated cultural effects on psychological processes. For example, we can separate the effect of holding particular value priorities oneself from the effect of living in a context where certain values are normatively prioritised. In an increasing number of studies, researchers are examining the effects of nation-level variables in such multi-level models, but relatively few have so far also controlled for the equivalent individual-level processes. In Chapter 7, we shall discuss some recent studies that have used such multi-level approaches.

Moderators are either categorical (gender, race, class, nation, etc.) or continuously measured variables (e.g., personality or values) that affect the direction and/or strength of the relationship between a predictor and criterion. In a correlational framework, the moderator is the third variable that affects the zero-order correlation between two variables, whereas in an ANOVA framework, they are represented as interactions between two independent variables. Moderator effects therefore refer to interactions between the variables used to predict an outcome.

This method has much potential for moving our understanding forward, especially if we can separate the effects of individual-level processes from those at the aggregated national level. Figure 4.5 shows the logic of such a multi-level model schematically. One further feature of multi-level models is that they can

also examine the strength of relationships, i.e., explore psychological processes, and whether these relationships differ across different cultural contexts. Variables that change the relationship between two other variables are called **moderators**. In later chapters, we will discuss a number of other studies that have identified such moderators in cross-cultural research. This is an exciting avenue for further research.

Unpackaging of culture is the process that explains why differences emerge between two or more cultural groups based on an explicit test of the psychological mediators of the observed cultural differences.

Mediators are variables that account for the relationship between an independent or predictor variable and the dependent or criterion variable. In psychological terms, they often explain how the external context takes on an internal psychological significance. Mediators more generally imply causal theoretical processes and explain how and why effects occur.

The Unpackaging of Culture

Returning to the level of the functioning individual, how can we study cultural processes if we have only a few samples and therefore cannot use multi-level modeling? What components of our outcome are cultural and what can be explained by other processes? Here, an old idea becomes important, the idea of **unpackaging** culture. Whiting and Whiting (1975) orchestrated a large ethnographic study of child development among six communities: a New England Baptist community; a Philippine *barrio*; an Okinawan village; an Indian village in Mexico; a northern Indian caste group; and a rural tribal group in Kenya. Observing substantial differences in psychological processes, socialization and child-rearing patterns, they reasoned that there

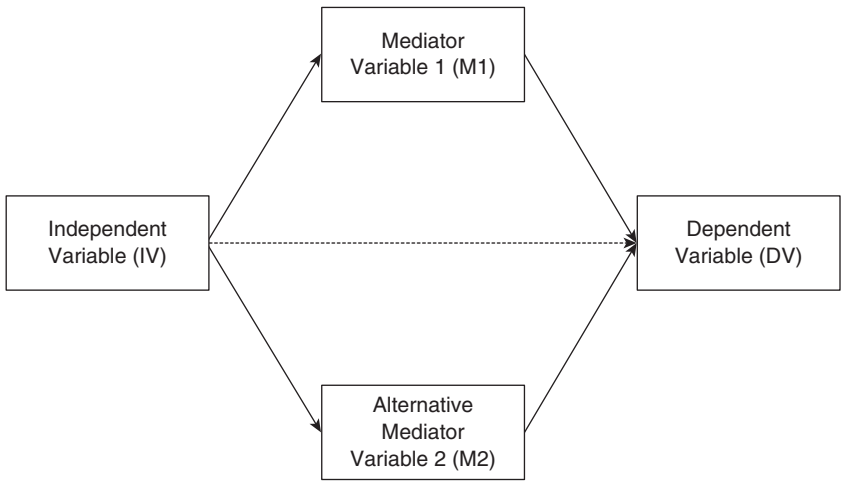

Figure 4.6 Example of a Mediator Relationship, with Two Potential Mediators

must be specific contextual variables that could explain the differences found, linking ecological constraints faced by these communities to psychological processes via adaptive socialisation practices. For instance, they compared the activities of children from the same families, some of whom were living in cities and others in villages. They also compared families in which young boys helped with baby-tending, with those in which girls did the helping. In ways such as these, the variability in observed behaviours could be broken down and associated with specific hypothesised mediators.

Hence, unpackaging studies are extensions of basic cross-cultural comparisons in which the active ingredient presumed to cause the observed differences in psychological processes is directly measured and explicitly tested for its efficacy in explaining the outcome. In the contemporary methodological literature, this process variable is called a **mediator**. Alternative names for unpackaging include 'linkage studies' (Matsumoto & Yoo, 2006), 'mediation studies' (Kirkman, Lowe & Gibson, 2006) or 'covariate studies/strategies' (Leung & van de Vijver, 2008).

Figure 4.6 shows a graphical representation of mediation. Most current mediation studies are conducted within a single nation, but we are especially interested here in those that have sampled several nations. Studies of this type provide a bridge between cross-cultural psychology and mainstream social/personality psychology, because they show that cross-cultural variations can be explained in terms drawn from both perspectives, without resorting to ill-defined constructs such as nationality (Brockner, 2003). For example, Tinsley (2001) found that differences in the conflict management strategies of German, Japanese and US managers were completely mediated by the values held by members of these cultural groups, and Felfe, Yan and Six (2008) reported that individuals' scores on a 'collectivism' scale mediated differences in organisational commitment across samples of Romanian, German and Chinese employees. Further studies testing mediation are discussed in Chapter 7.

In an ideal test of mediation, the researcher tests whether other relevant variables that are *not* related to the hypothesis also yield mediation effects. This provides greater certainty in establishing exactly what causes the results that are obtained. For instance, in a study discussed in Chapter 10, Y. Chen, Brockner, and Katz (1996) showed that a measure of individual-collective primacy mediated the intergroup effects that they had predicted and found. They then tested whether six other measures derived from the concept of individualism-collectivism also mediated these effects, and found that they did not. Studies of this kind help to clarify the loose and varied ways in which the psychological aspects of individualism and collectivism have been employed by different authors.

Another approach is illustrated by the work of Van de Vliert and Smith (2004), who tested predictions derived from Van de Vliert's theory of the effects of climate and wealth, which was discussed in Chapter 3. Across 76 nations, the climate and wealth predictions of variation in leader style were upheld. These authors then evaluated the distinctiveness of their findings by testing whether they were mediated by Hofstede's measures of uncertainty avoidance and power distance. No mediations were found, even though these Hofstede scores did each correlate

significantly with leader style. This enabled Van de Vliert to sustain his claim that he had identified a distinctive predictor of cultural differences.

In summary, unpackaging has two inter-related features: identification of the theoretical factors or processes that may cause cultural differences in psychological outcomes of interest, and an explicit empirical test of the proposed processes leading to these outcomes (see Poortinga & Van de Vijver, 1987; Leung & Van de Vijver, 2008; Fischer, 2009, for more technical explanations).

META-ANALYSIS: A USEFUL TOOL FOR INTEGRATING RESEARCH

Thus far we have focused on methods that sampled individuals in different locations. In summarising the results of intensively researched fields with many independent studies, meta-analysis can be a fruitful tool with which to study cultural phenomena. A meta-analysis is a set of quantitative techniques that aggregates results across many (typically published) studies. In other words, it is an analysis of previous analyses. Meta-analysis can be conducted so long as you have an effect size, that is, a numerical measure of the expression of a psychological characteristic (a mean or frequency), or the strength of an association between two psychological constructs (typically a correlation), or the mean difference of a psychological characteristic between two or more groups.

The results reported from previous studies may need to be converted, so that they can be compared directly. For instance, they could be weighted by sample size, so as to ensure that a study with 1,000 participants receives more weight than a study with only ten participants in the conclusions. Each result could also be adjusted for criteria of quality, so as to give more weight to studies that were well designed compared to those that had many flaws. Once we have a combined mean effect, we can test whether the studies are homogeneous, that is whether they show the same effect. If there are differences between studies, we can search for moderator variables that can explain such variability. Meta-analysis is a sensitive tool for summarising and synthesising past research, testing novel hypotheses, and identifying gaps in the literature.

Throughout the book, we present various meta-analyses to summarise past research. Traditionally, cross-cultural meta-analysis has focused on cultural differences, requiring a comparison of at least two or more samples. With the increasing number of studies in single nations, we can also combine single-nation studies for cross-cultural meta-analyses. For example, we could examine whether the means of a variable vary systematically with nation-level indicators or we could examine whether two variables correlate differently in different ecological contexts. Meta-analysis in this context is essentially a multi-level analysis, allowing us to understand how psychological processes vary across samples and nations (e.g., Fischer & Mansell, 2009; Fischer & Boer, 2011; Boer & Fischer, 2013). Meta-analyses therefore provide another option for researchers who wish to unpackage the differences between nations.

CONCLUSION

Studying the Ik

If we were to understand the psychological processes that led Adupa's parents to starve their daughter, this chapter has provided many options. We could think of the value of children as a crucial variable that has some role to play in the process that Turnbull observed (see cross-cultural research into the value of children in Chapter 8). Alternatively, we could focus on the effect of resource availability on parenting behaviours. What other variables could you think of that might be applicable in this case?

We could undertake an ethnographic study and conduct interviews with mothers and fathers in a number of traditional and modern societies. We could conduct some simple experiments in which we give vignettes or scenarios to parents and ask how they might behave in a situation like that. We could use some form of economic game in which money or some other resource needs to be divided between themselves and their children (or between their children). We could develop tests that measure moral concerns or devise a questionnaire that captures the psychological value of children to their parents. We could adapt existing measures or develop new ones that fit each of the cultural contexts of our study. For example, does the number of children the parents already have make a difference? What about their gender, or their ages?

If we had only a few samples, we could use unpackaging at the individual level to see what psychological processes are implied. If we had studied more than ten samples from around the world, we could then use multi-level models to understand how features of the context affect such decisions. Obviously, we would need to consider carefully the samples that we should recruit and how we might control for any other variables that could influence our results. Alternatively, we might want to include such variables as additional independent variables, to further understand parental decisions in resource allocation to their children. For example, we could treat access to food as an independent variable in its own right. What other variables would we need to control or include in our design?

FURTHER READING

1 He, J., & Van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture.* Available at http://scholarworks.gvsu.edu/*orpc*
2 Matsumoto, D., & Van de Vijver, F. (2011). *Cross-cultural research methods in psychology*. Cambridge: Cambridge University Press.
3 Valsiner, J. (2003). Culture and its transfer: Ways of creating general knowledge through the study of cultural particulars. *Online Readings in Psychology and Culture.* Available at http://scholarworks.gvsu.edu/*orpc*
4 Van de Vijver, F. (2009). Types of comparative studies in cross-cultural psychology. *Online Readings in Psychology and Culture.* Available at http://scholarworks.gvsu.edu/*orpc*
5 Van Hemert, D.A. (2003). Cross-cultural meta-analyses. *Online Readings in Psychology and Culture.* Available at http://scholarworks.gvsu.edu/*orpc*

STUDY QUESTIONS

1 Should different research methods be used in studying different cultures or are there methods that can usefully be applied in all cultural contexts? Give examples.
2 Which research method holds the greatest promise for advancing our understanding of cross-cultural issues: fieldwork, psychometric tests and surveys, or experimentation? Explain why.
3 Select any one of the five guidelines for cross-cultural research and explain why it is important in the study of social psychology across cultures.
4 Explain the difference between mediation and moderation.