East Asian wisdom and relativity

Inter-ocular testing of Schwartz values from WVS with extension of the ReVaMB model

Chris Baumann
*Faculty of Business and Economics, Macquarie University,*
*Sydney, Australia and*
*College of Business Administration, Seoul National University (SNU),
Seoul, Republic of Korea*

Hume Winzar
*Faculty of Business and Economics, Macquarie University,*
*Sydney, Australia, and*

Tony Fang
*Stockholm Business School, Stockholm University,*
*Stockholm, Sweden*

Abstract

**Purpose** – The purpose of this paper is threefold. First, the paper demonstrates how inter-ocular testing (looking at the data) of Schwartz values from world values study (WVS) provides a surprisingly different picture to what the authors would expect from traditional mean comparison testing (*t*-tests, analysis of variance (ANOVA)). Second, the authors suggest that the ReVaMB model can be applied to an East Asian philosophical perspective. Relativity, the authors argue, is a factor when East Asian wisdom, philosophies and ideologies (Confucianism, Taoism, Buddhism and Legalism) “drive” outcomes such as work ethic. Third, the paper serves as an editorial to a special issue in *CCSM* on East Asian wisdom and its impact on business culture and performance in a cross-cultural context. Common themes are Yin Yang, how different cultures deal with paradox, and Zhong Yong, with accompanying concerns of how to conceptualise and deal with balance of opposites.

**Design/methodology/approach** – The authors adopted ten variables of the Schwartz values scales used in the WVS and subjected them to principle components analysis to reduce the number of variables. The authors found a two-factor solution: one relating to personal material success and adventure and excitement; another relating to success and personal recognition. The authors labelled these factors as Altruism and Hedonism. The analysis is based on an overall sample of 84,692 respondents in 60 countries. In addition to traditional statistical testing, the authors conduct inter-ocular testing. The authors also suggest that the ReVaMB model can be applied to East Asian wisdom.

**Findings** – Three recommendations help to arrive at more accurate conclusions when comparing groups:

- The authors recommend to aspire to “consistent look and statistic”. If the data distribution does not agree with the statistics, then the researcher should take a closer look. To avoid misinterpreting statistics and other analysis, the authors recommend inter-ocular testing, i.e. eyeballing data in a scientific fashion.

**Originality/value** – The authors introduce two concepts: the “inter-ocular test”, which simply means to “look at your data”, and the Chinese word, 错觉 (Cuòjué) which roughly translates to “illusion”, “wrong impression”, or “misconception”. This study argues against accepting simplistic averages for data analysis. The authors provide evidence that an inter-ocular test provides a more comprehensive picture of data when comparing groups rather than simply relying on traditional statistical mean comparison testing. The “word of caution” is to avoid premature conclusions on group comparisons with statistical testing alone.

An important point to make is a word of big thanks to Professor Rosalie L. Tung, Editor-in-Chief of *CCSM*. Rosalie is the first author’s long-standing mentor and friend, and has been instrumental in Chris’ career and publication Mantra. Chris concludes that from Rosalie, one learns true excellence. All three authors are most grateful for Professor Tung’s guidance on this special issue.
The authors also propose an extension of the original ReVaMB model from a Confucian orientation to a broad East Asian philosophical perspective. Culture does determine attitudes and behaviour which in turn contribute to the shaping of cultures, depending on situation, context, location and time. The “context” for a situation to occur should be tested as moderators, for example, between East Asian wisdom (Confucianism, Taoism, Buddhism and Legalism) and behavioural or attitudinal dimensions such as work ethic.

**Keywords** Confucianism, Common-language effect (CLE), Inter-ocular testing, ReVaMB model, World values survey (WVS), Ying Yang, Cuòjué

**Paper type** Research paper

1. **Introduction**

The purpose of this paper is threefold. First, we demonstrate how inter-ocular testing of Schwartz values from world values study (WVS) provides a surprisingly different picture to what we would expect from traditional mean comparison testing such as *t*-tests and ANOVA. As such our paper will later result in a “word of caution” for (often) premature conclusions about culture and ethnic groups being different – or not. Second, this paper also serves as an editorial to a special issue in CCSM on East Asian Wisdom and its impact on business culture and performance in a cross-cultural context. Third, we recommend the ReVaMB model previously introduced by Baumann and Winzar (2017) with a lens of Confucianism originally, to a broader East Asian philosophical perspective. After all, how culture “drives” attitudes and behaviour which in turn “shapes” the culture often depends on relativity, namely, the situation, context, location and time. Relativity, we argue, is also a key factor when we look at East Asian wisdom and how Confucianism, Taoism, Buddhism and Legalism, to name a few key East Asian philosophies and ideologies, determine outcomes such as work ethic. In statistical “speak”, circumstances should be incorporated into regression/SEM/PLS modelling as moderators.

Given the long history of East Asian culture to emerge from the combination of Confucianism and Taoism in China, spreading to Korea, Japan and Vietnam, it is not necessarily easy to “capture” East Asian culture and wisdom. Confucianism is also “joined” by and/or intertwined with Taoism, Buddhism and Legalism, and all have experienced changes over time while spreading geographically. Taoism and Confucianism emerged almost at the same time with Confucius a few years younger than Lao Tzu. In Chinese literature, it is written that Confucius visited Lao Tzu to ask the meaning of Tao. Confucianism (Yang) and Taoism (Yin) are equally powerful and influential in the Chinese psyche but in different ways. The former is more visible and often officially adopted than the latter. At the same time, East Asia experienced other cultural, economic and ideological influences, ranging from the Soviet Union’s Communism to Western capitalism; to make things more complicated, there has also been the influx of major world religions (e.g. Christianity, Buddhism). As such we are looking at not just East Asian Wisdom, but East Asian Wisdoms! When we consider the different histories and cultural evolutions that have emerged in each country and region, we see a fascinating combination of Westernisation, Modernisation, Confucianism, Taoism, Buddhism and Legalism. Here, we refer to Westernisation as the cultural influence of predominantly European and, most recently, American cultural artefacts – music, food, entertainment, movies and TV, language pronunciation and slang, fashion and so on. In contrast, Modernisation is the process of using the most recent ideas, technology, infrastructure and methods so that goods and behaviour are, or appear to be, more modern (Hill, 2000). Confucianism, Taoism, and Buddhism are the three main Chinese philosophical thoughts (Fang, 1999). Today’s China is westernised compared with China under Mao. But, today’s China is not westernised as much as it is modernised. The following table crudely illustrates our view of the dominant factors influencing life and business in just five country/regions of North-East Asia: Urban China, Rural China, Taiwanese China, South Korea and Japan (Table I).
This table is limited, obviously, religion and spiritualism could be added. And we have left out most other East Asian countries, and we could argue about inclusion of Rural and Urban China. But the point remains, each of these region/countries has arrived at its current state of culture and economic development through different histories and influences. And each is enjoying extraordinary success. The combinations of ingredients – maintenance and adaptation of traditional values, adoption and adaptation of Western and/or Modern technologies and systems – have produced rapid improvements in the welfare of citizens, and global influence.

East Asia has also seen dramatic economic progression and development. For example, directly after Second World War all countries in the region were of a low-level development, but are now advanced economies with leading innovation in the car sector (including environmentally friendly cars), electronics and manufacturing industries, and more recently also in services (hospitality, transportation) and experiences (theme parks, beauty and plastic surgery; entertainment), Artificial Intelligence (AI), and the list goes on. East Asia has seen massive modernisation, but to what degree has that trend been coupled with westernisation? In sum, modernisation may have occurred with different degrees of westernisation for different countries, and within different countries (Tu, 2014). For example, Taiwan is very traditional Chinese in many ways, preserving Confucian Chinese tradition at least at the level of Mainland China.

In education, Chinese Taiwan actually teaches Confucianism in schools as part of the curriculum, whereas on the Chinese Mainland, Confucianism is generally offered as an elective subject. Regardless, in China, Chinese Taiwan, Japan and Korea, the teaching style remained largely Confucian, i.e. with strict discipline, an emphasis on respect, and even more so on human betterment (Baumann and Krskova, 2016). East Asia managed to massively modernise its school curriculum and infrastructure, yet maintaining a strong Confucian pedagogic approach, ultimately resulting in the world’s strongest academic performance, at least according to PISA. In Asia, strict discipline in schools and a focus on academic performance in educational institutions have been found as determinants of work ethic, but not so in Western countries (Baumann et al., 2016). Such a finding may be, to some extent at least, an outcome of students being viewed – and treated – as learners under Confucianism, in contrast to – more and more – as customers in the “West”.

The Yin Yang thinking is a combination of three principles, i.e., holism, dynamism and duality (Fang, 2012; Li, 2016). The Yin Yang thinking is “either-and” (Li, 2016) which embraces but also goes beyond the focused “average”. Yin Yang is about the point and area; whereas “average” is only about the point. The Yin Yang thinking allows East Asia to embrace opposing approaches, such as a traditional yet very modern approach to management, education and running societies at large. While this is fascinating, it makes for a challenging task to measure, conceptualise and test East Asian philosophy and wisdom in a scholarly, empirical fashion (Fang et al., 2017; see also Pauluzzo, Guarda, De Pretto and Fang in this special issue). Li (2016) highlights the value of Yin Yang thinking as indigenous

<table>
<thead>
<tr>
<th>Westernisation</th>
<th>China (Urban)</th>
<th>China (Rural)</th>
<th>Taiwanese</th>
<th>Korea</th>
<th>Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modernisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confucianism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taoism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buddhism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legalism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Philosophical variation in the Confucian orbit

Note: The more intense the shading, the greater the extent to which that particular trait is manifested in that country/region
epistemological system from the East to understand different contexts. Straightforward either-or situations are best dealt with using Western-style absolute thinking, whereas paradoxical, either-and, situations require acceptance, even encouragement, of Eastern-style thought (Zhu et al., 2016). In many ways what Western researchers consider new has been understood for centuries in East-Asia. For example, Chen (2016) discusses at length the similarities between Eastern, specifically Chinese, philosophy and Western competitive dynamics. He points out that co-opetition, a relatively new concept in Western managerial thought, is normal in Eastern classical literature. Contrasting Asia vis-à-vis the “West” on competitiveness, competitive attitude, willingness to serve and speed have been identified as drivers of workforce performance in Asia, contrary to Western countries where these three motivational drivers did not have strong explanatory power (Baumann et al., 2016). In simple terms, “East and West” do seem to differ on a not insignificant number of dimensions in relation to motivation, thinking and performance, to name a few.

Yin Yang sees value in combined perspectives – harmony between conditions rather than conflict between opposing phenomena. Such an approach attempts to incorporate the whole rather than simplified abstractions of a phenomenon. At the risk of oversimplification, for example, Western attempts to measure and analyse aspects of culture have relied on simple averages for a cohort. We believe that is a serious mistake.

2. Countries are not their average

Tsui (2016) “explains the two primary types of values relevant for science: epistemic – norms and standards to ensure good science – and social – criteria not relevant for discovering the truth of knowledge but may influence decisions related to science especially in evaluating the cost of wrongful conclusions from the research evidence”. The focus only on averages is not only poor epistemology but risks dangerous conclusions when we make business decisions or government policy based on such crude measures. Implicit in Tsui’s argument is a profoundly Confucian view of science – that it is important and valuable to understand the nature of the world, but the implications for society are equally as important. Under Confucian philosophy, it is a duty of each person to improve oneself and work for the betterment of those around one. Tsui reminds us of recent work by Birkinshaw et al. (2014), who criticised the “scientistic” approaches of some research – the rituals and appearance of science that does little to produce new knowledge. Healey argues, “the bulk of management research subsists on analysing regularities in statistical surrogates and drawing inferences from self-reports that are often twice removed from the action. The benefits of close contact are many; not least insight, inspiration, curiosity and ecological validity” (Birkinshaw et al., 2014, p. 47). Arguably, the same can be said for much of the cross-cultural research we have consulted. Ironically, some “closer contact” can be found in the usual statistical methods, by considering the range of our measures, rather than their averages, and the context in which those measures are made.

2.1 错觉 (Cuòjué)

For want of a better term, we introduce the Chinese word 错觉 (Cuòjué), which roughly translates to “illusion” or “wrong impression”. We argue that, too often, we read arguments about culture or organisational behaviour based on averages of the individuals within those cohorts. The average of a group is not the group, and it certainly is not any individual in the group. Modelling often is a simplistic replication of earlier work that may not consider the context of the problem, or without thinking about the meanings of scales and measures. For example, Baumann et al. (2011, p. 252) refer to Jarvis et al. (2003), and observe that “[...] some 96 per cent of SEM studies published in leading journals used reflective measures; while only 4 per cent used formative measures. Perhaps more alarmingly, they claim that 28 per cent of the reflective models should more correctly have used formative measures”.

East Asian wisdom and inter-ocular testing
Leading international business journals, e.g. *Cross Cultural and Strategic Management (CCSM)* and the *Journal of International Business Studies (JIBS)*, now require testing, or at least discussion of, *Common-Method variance* (CMV), in manuscript submissions; CMV being the spurious “variance that is attributable to the measurement method rather than to the constructs the measures are assumed to represent” (Yüksel, 2017, p. 377). Researchers often face two challenges when collecting primary data: First, collect one set of survey data, but then face the problem that CMV could be present (e.g. “inflated” or “deflated” responses/values in cross-cultural research). That would require subsequent testing for possible CMV effects, principally applying three different techniques: harman single factor, common latent factor, and common marker variable (see, e.g. Eichhorn, 2014; Williams and Anderson, 1994; Williams *et al.*, 2010). Second and alternatively, in order to avoid CMV by research design, independent/moderating/mediating variables would ideally be separately and independently measured from dependent variables; see, e.g. Baumann and Harvey (2018).

Hofstede (2015, p. 556) notes that the use of agent-based models pioneered by Schelling (1971) and Sakoda (1971) “has led to the fields of computational social science, artificial economics and social simulation, fields that investigate the links between micro-level behaviour of agents – often individuals – and resulting macro-level patterns. Culture is such a pattern, and could thus be studied using these techniques”. In other words, Hofstede’ son, Gert Jan, has consciously stepped away from defining culture as simply the aggregation of individual-level responses to questions, as sometimes implied by his father, Geert Hofstede (1980), and too often taken as a reified truth by some scholars. Instead, we should look at culture as patterns of behaviour that emerge in a non-linear and non-aggregate way from individual values and behaviours, in an interactive social environment.

Confusion about level of measurement and the effects of average scores is often subtle and pervasive. Applying the average of the group to the individuals in the group is known as the Ecological Fallacy (McSweeney, 2002, 2013; Winzar, 2015). Misunderstanding of the Ecological Fallacy has led to some creative analyses. For example, Andreassi *et al.* (2014, p. 70) acknowledge in their Limitations section that “culture even at a national level is an aggregate measure of individual values and is not generalizable as the measure of each individual’s perception of these cultural dimensions”. This statement was made after the researchers had applied averages to the individuals. In another example, Song *et al.* (2017) mixed different units of measurement by combining national-level constructs with individual-level evaluations. Individuals were questioned about a national-level construct, then averaged across all individuals, and then applied back to individuals. This was a rather “tricky” analysis that risked both the ecological fallacy and the atomistic fallacy, inferring the behaviour of a whole group based on the behaviour of a few individuals (Lieberson, 1991). To further explain this point, we use WVS as an example in the following section.

### 3. Re-evaluating the WVS

Ten variables designed to capture aspects of the Schwartz values scales were used in the WVS Wave 6 (WVS, 2016) (variables V71:V79). We subjected them to principle components with varimax rotation to reduce the number of variables to a more manageable number and to remove correlation among constructs. After removing those with communalities of less than 0.40, and ensuring that the solution was consistent across all countries, we found the two-factor solution in Table II. We can see that four variables load highly on Factor no. 1, each relating to such values as helping other people in the community, environmental concerns, tradition and behaving with propriety. Two variables loaded highly on Factor no. 2, relating to personal material success and adventure and excitement. Another variable, relating to success and personal recognition, loaded almost equally on both factors. A three-factor
solution did not resolve this loading issue. We have labelled these factors, somewhat subjectively, as Altruism and Hedonism, respectively. We extracted then two orthogonal factors and applied them to all respondents in the WVS (2016) data set. This provided us with two uncorrelated variables with which to examine all respondents in the data set within each country and cultural grouping. The variables were standardised across all 84,692 respondents in 60 countries, so that scores for both Altruism and Hedonism have Mean of zero and Standard Deviation of one. Thus, about 95 per cent of cases have scores ranging between $-2$ and $+2$, and 99.7 per cent of cases have scores in the range $-3$ to $+3$.

4. Data presentation and interpretation

Analysis of statistical data and comparison of groups in this journal and many others traditionally is dominated by null-hypothesis significance testing (NHST) – $t$-tests, ANOVA, etc. – with much concern for the $p$-value. A $p$-value is supposed to be an estimate of the probability of getting a given result when there is really no relationship or difference. We want to avoid type 1 error (hanging an innocent man). The $p$-value says nothing about type 2 error (releasing a guilty man). Additionally, the $p$-value is not a reliable test when there really is a difference between two groups (Cumming, 2012). For some scholars it can be an obsession: $p > 0.05$ means failure; $p < 0.05$ means celebrate and submit for publication. The American Statistical Association has long campaigned against such naive reliance on $p$-values, with some success (Wasserstein and Lazar, 2016). A recent editorial in *Journal of International Business Studies (JIBS)* (Meyer et al., 2017), raised concerns about such reliance on $p$-values, and offered ten concrete guidelines for avoiding false positives (type 1 errors). We humbly offer one additional guideline: look at the data.

Our discussion here is prompted by recent reminders that researchers should look at their data, for example, the datasaurus, before undertaking formal data analysis (Matejka and Fitzmaurice, 2017). Some years ago, the first and second authors’ friend and

<table>
<thead>
<tr>
<th>Schwartz variable</th>
<th>Factor no. 1 (Altruism)</th>
<th>Factor no. 2 (Hedonism)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Looking after the environment is important to this person; to care for nature and save life resources (V78)</td>
<td>0.775</td>
<td></td>
</tr>
<tr>
<td>Help others, mean of two variables: it is important to this person to do something for the good of society (V74) and It is important to help people living nearby; to care for their needs (V74B)</td>
<td>0.735</td>
<td>0.170</td>
</tr>
<tr>
<td>Tradition is important to this person; to follow the customs handed down by one's religion or family (V79)</td>
<td>0.716</td>
<td></td>
</tr>
<tr>
<td>It is important to this person to always behave properly; to avoid doing anything people would say is wrong (V77)</td>
<td>0.710</td>
<td>0.102</td>
</tr>
<tr>
<td>It is important to this person to be rich; to have a lot of money and expensive things (V71)</td>
<td>0.787</td>
<td></td>
</tr>
<tr>
<td>Adventure and taking risks are important to this person; to have an exciting life (V76)</td>
<td>0.766</td>
<td></td>
</tr>
<tr>
<td>Being very successful is important to this person; to have people recognise one's achievements (V75)</td>
<td>0.411</td>
<td>0.650</td>
</tr>
</tbody>
</table>

Notes: Questions relate to Schwartz's values inventory. Respondents read the questions and marked a six-point rating scale ranging from 0 (not at all like me) to 5 (very much like me). Orthogonal factors were extracted based on the above factor loadings, and standardised so that the Altruism score and the Hedonism score had Mean of zero and Standard Deviation of one across all 84,692 people in all 60 countries; Factor loadings less than 0.1 hidden for ease of reading; V – denotes the variable number in the World Values Survey (6); World Values Survey (2016). World Values Survey wave 6 (2010-2014), official aggregate. Asep/JDS, Madrid Spain, World Values Survey Association. www.worldvaluessurvey.org

Table II. Exploratory factor analysis of WVS Schwartz variables
mentor Professor Lester W. Johnson, coined the term “inter-ocular test” to encourage his students to always “eye-ball the data”. We offer the following breakdown of the standard approach to aggregating survey data. We can avoid Cuòjué (self-deception, illusion and misunderstanding) by using the “inter-ocular test”. When comparing survey results from two countries, we should first graph the data: see the extent that the data overlap or not, and consider comparative ranges and values.

Let us consider first the usual approach to cultural comparisons – averages, $t$-tests and ANOVA. Figure 1 presents the average scores on both Altruism and Hedonism across different national cultural groupings as suggested by Minkov (2011). With Figure 1 we can construct a very plausible story that reinforces a colonial world view and much of what has passed as scholarship in Western literature. Those in the upper-right of the graph might be inferred as the “good” people: they enjoy life, and they want others to enjoy life also. And of course, at the other end of the graph are the “bad” people, those who do not enjoy life and do not care for you to either.

Moving on to Figure 2, showing all countries in the WVS 6 data set, the story is not as clear, but we could probably make up a plausible story just the same. Average scores for all the Anglo and Western European countries are still in the upper-right corner, and we could argue that some Eastern European countries are more “developed”, socially or economically, to justify their placement. Similarly, the same “good/bad” subtext can be inferred for most other countries. The analysis usually ends here. The reader is probably aware of several books and journal articles that present persuasive maps of country and cultural centroids such as Figures 1 and 2 where it is implicit that one country, or one cultural group, is “better” than another on some dimension. We are ignoring, of course, the underlying assumption in this and all similar studies, that all cultures interpret the questions the same way, and use the six-point response scales the same way. For now, let us consider the scale of all these means. The data are normalised, meaning that the mean among all (84,692) respondents is zero and standard deviation is one. This rescaling permits easy comparison across all individuals and groups using the same scales of measurement. Values across all

![Figure 1. Average personal values across cultural groups](image)
people range from $-3.0$ to $-3.0$ (six standard deviations). Figures 1 and 2 show mean scores for cultural groups range from $-0.4$ to $+0.4$ (only 0.8 of a standard deviation). The averages are really all bunched together, as we can see in Figure 3, which presents the centroids in the context of the full range of the data.

Now let us consider Figures 4 and 5. In each figure we have attempted to present, in Column 1, a bar-graph showing only the mean scores for each variable in each country within its Minkov cultural grouping. It is easy to draw the conclusion that some countries are high, and others are low. But the second data column suggests a subtly different story. In Column 2 we see each respondent graphed under a boxplot showing the mean,
plus interquartile range and 95 per cent range. The scales of the two data columns are the same so that the centre of the boxplot corresponds with the end of the bar-graph. Note the amount of overlap in the data among any two countries. Generally, there is much more overlap than there are separate observations. Now let us ask the implicit question more openly [...] Is any one country really that much different from another? Some are, but most are not.

5. Comparisons of country pairs
Historically, when comparing two or more groups, we test whether the means are different, using a *t*-test or ANOVA and the standard measure of difference is the *p*-value, which defines the level of statistical significance. Unfortunately, with large sample sizes, almost everything is “statistically significant”. With large sample sizes, statistical significance is rarely meaningful. When William Sealy Gosset published what became the *t*-test, he had sample sizes of literally a handful in mind: his table of “significant” values represented samples of between four and ten only (Gosset, 1908). Further, what Gosset called “significant” did not mean “important”. He simply meant it to be a call to look closer.
When samples are in the thousands, then miniscule differences in sample means appear to be “significant”. If we want to determine how much one group is different from another then measures of effect size are more appropriate.

5.1 Effect size
Effect size can be seen “as a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest” (Kelley and Preacher, 2012, p. 137). When comparing groups, effect size is a measure of how strong is a difference. We offer three measures: the common-language effect size (CLE), Cohen’s D, and $\eta^2$.

1) CLE: CLE is an intuitively simple statistic suggested by McGraw and Wong (1992). CLE is a measure of the proportion of times one randomly-selected member of one group scores higher (or lower) than one randomly selected member of another group. For example, if we have ten men and ten women we can compare the heights of each man with each woman – 100 paired comparisons. Of these we might find that 80 of those 100 comparisons resulted in the man being taller than the woman. So, the CLE is 0.8, or 80 per cent. CLE is robust against outliers and
sample size, and it is relatively easy to understand for decision-makers who are not confident or familiar with statistical terminology (Brooks et al., 2014).

(2) **Cohen’s D**: Cohen’s D is an effect size used to indicate the standardised difference between two means Cohen (2013). It is the number of standard deviations distance between the two means. The lower bound for D is zero, when the two means are equal, but there is no upper bound. Sawilowsky (2009) suggests rules of thumb for effect sizes to define D (0.01) = very small, D (0.2) = small, D (0.5) = medium, D (0.8) = large, D (1.2) = very large, and D (2.0) = huge. Of course, Cohen warns against using such rules of thumb as gospel. For example, if all comparisons between two groups in a study are extremely small then a slightly larger statistic might be cause for further investigation.

(3) **\(\eta^2\)**: \(\eta^2\) is analogous to \(r^2\); the proportion of variance shared by two variables. In the case of two independent samples, it is the amount of variation in the data that can be attributed to the grouping variable. For example, if comparing two groups, if \(\eta^2\) is 0.3, then 30 per cent of the variation in the variable of interest is accounted for by knowing the two groups.

With effect size in mind, it is worth reconsidering our historical perspectives on group differences based only on statistical means.

Figure 6 presents summary information for different pairs of countries selected from the WVS 6 data set. We have presented a density plot graph for each country so that the reader can see the distribution of scores for both variables, Altruism and Hedonism. Within each graph we include the following statistics:

- **t-test**: a regular independent samples t-test of whether the two means are different from each other, plus degrees of freedom, and standard two-tailed “significance level” as measured by the \(p\)-value.
- **\(\eta^2\)** (“quasi \(R^2\)”), to show the proportion of variation explained by recognising that the data come from two different groups.
- **Cohen’s D**, the number of standard deviations separating the two means.
- **CLE**, presented as the proportion of times that members of one group scored higher than members of the other group. In this case, we took each respondent for one country and compared it with each case in the other country. For example, comparing Taiwan with China, each of the 2453 Taiwanese respondents was compared with each one of the 2612 Chinese respondents – 6,407,236 paired comparisons.

Collectively, are the Taiwanese Chinese different from Mainland Chinese people, in other words, what is the real difference between the Mainland Chinese and Taiwan Chinese? A standard t-test suggests that they are quite different on both Altruism and Hedonism, with \(t\)-values of 4.7 and 24.0, respectively. Figure 6 also shows that there is a great deal of overlap between Taiwan and China on Altruism. CLE scores are 55/45 per cent, and \(\eta^2\) is less than 1 per cent. The means, as measured by Cohen’s D, are only 0.17 standard deviations apart. Overall, are the Mainland Chinese and Taiwan Chinese really different with respect to Altruism? We suggest not. Hedonism may be a different issue for Taiwan and China. CLE scores are 73/27 per cent in favour of Taiwan, and \(\eta^2\) is over 14 per cent. We can conclude then that, generally, the Taiwan Chinese enjoy life more than the Mainland Chinese.

Comparisons of Singapore with China give similar results: almost identical distributions for Altruism but somewhat different for Hedonism. Here, the Chinese seem to be more hedonistic than the Singaporeans.

What is sufficient to conclude there is a difference? That is a judgement call. Like all guidelines for what is worthy of investigation or highlighting (including the notion that
Figure 6. Paired distributions of Altruism and Hedonism for selected countries (continued)
Figure 6. (continued)
A p-value should be less than 0.05 to be called “significant”) the decision should be based on the needs of the research and relative to other measures. We felt that a CLE split of 60/40 per cent was the minimum difference to make any generalisations about two countries. 70/30 per cent is worth talking about. With that crude rule of thumb, we can see...
that there are very few big differences across the country pairs. We have included several
more country pairs for the reader to examine.

Examination of the different graphs in Figure 6 highlights some interesting cases where
the data distributions match with the traditional statistical inference, the t-test. These are
summarised in Table III. When frequency plots or density plots are consistent with the
t-test, then we are satisfied. But when the appearance of density plots is not consistent with
the t-test, then we risk falling into what we call the Cuòjué trap. We use a framework similar
to types 1 and 2 errors. Type 1 error is a false positive, rejecting the null-hypothesis when
the null is true, or simply “hanging the innocent man”. Type 2 error is a false negative,
accepting the null when it is really false, or simply “releasing the guilty man”. We suggest
two types of Cuòjué or confusion, illusion, misunderstanding when examining a p-value and
comparing it with the data distributions under consideration. This applies to t-tests,
ANOVA, and other traditional tests involving a p-value:

- **Type 1 Cuòjué** occurs when we calculate a statistically significant p-value, but the
data distributions appear very similar. Type 1 Cuòjué occurs when we test relatively
large samples, as we have in the WVS 6 data set. Type 1 Cuòjué means that, for
example, while a traditional t-test may suggest that two means are different, for all
practical purposes, there is no real difference in the two groups.

- **Type 2 Cuòjué** occurs when the data distributions look different, but the p-value
suggests that they are not. Type 2 Cuòjué is common with small samples.

Our analysis leads us to make some broad conclusions that we hope our colleagues in
international business, including cross-cultural researchers, might follow up on:

- Averages alone are rarely useful. The full range of values gives a much more
nuanced understanding of the data.

- Looking at (and thinking about) the data helps.

- These considerations may undermine much of what has passed for scholarship in our
area. This should be a concern for some researchers – accepting averages as
representative of a whole country or cultural group is to fall victim of the ecological
fallacy. It is a cause for more deep and thoughtful analysis.

The theoretical implications of this study contribute mainly to the debates on whether,
and how, culture “drives” behaviour, within the context of specific situations, without
discounting the very real probability of “alternative causal directions” (Baumann and
Winzar, 2016, p. 21). That is, modified behaviour of sufficient numbers of people may
affect group culture over time. But our paper makes a stronger contribution to
methodology, namely we provide strong evidence that simple means testing alone could
indeed lead to faulty conclusions, for example that a significant difference is found, but in

<table>
<thead>
<tr>
<th>Density plot (inter-ocular test)</th>
<th>t-test (“traditional” statistical inference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not statistically different</td>
<td>Statistically different</td>
</tr>
<tr>
<td>Similar distribution</td>
<td><strong>Consistent look and statistic</strong></td>
</tr>
<tr>
<td></td>
<td>Looks the same/no statistical difference</td>
</tr>
<tr>
<td></td>
<td>Requires high-level of overlap</td>
</tr>
<tr>
<td>Different distribution</td>
<td><strong>Type 1 Cuòjué</strong></td>
</tr>
<tr>
<td></td>
<td>Looks the same/statistically different</td>
</tr>
<tr>
<td></td>
<td>Common with large samples</td>
</tr>
<tr>
<td></td>
<td><strong>Consistent look and statistic</strong></td>
</tr>
<tr>
<td></td>
<td>Looks different/is statistically different</td>
</tr>
<tr>
<td></td>
<td>What we had like to see</td>
</tr>
</tbody>
</table>

| Table III. Avoiding Cuòjué: patterns of “inter-ocular” test and “traditional” statistical inference |
fact two groups are not different substantially in reality. Our paper provides a framework/approach how data can be plotted, and directly contrasted to mean comparisons to show the full picture.

Our contribution to methodology centres on the following three very practical recommendations:

1. **Type 1 Cuòjué and type 2 Cuòjué**: in our Table III we show patterns of “inter-ocular” test and “traditional” statistical inference. Naturally, our recommendation for researchers is to aspire “consistent look and statistic”, and to avoid the two types of Cuòjué we have identified. Our table should prove useful to identify the “danger zones”, and ultimately avoid them.

2. **Inter-ocular testing**: to avoid the above types 1 and 2 Cuòjué, we recommend inter-ocular testing in addition to traditional statistical testing, i.e. eyeballing the data in a scientific fashion. We provide specific examples in Figures 4 to 6 on how such inter-ocular testing could be executed.

3. **Common-language effect (CLE) size**: we recommend to test for CLE, and we further recommend a new rule of thumb, i.e. a split of 60/40 per cent as the minimum difference to make any generalisation about two countries (or cultural, ethnic or religious group; or gender; or any other categorisation); 70/30 per cent is worth considering. With our crude new rule of thumb, it is likely that a better differentiation between real and “not real” differences can be made.

In summary, we argue against accepting simplistic averages for data analysis. We provide empirical evidence that an inter-ocular test, looking at the data, provides a more comprehensive picture of data when comparing different countries, cultures or ethnic groups in contrast to simply relying on traditional statistical mean comparison testing (namely, $t$-tests, ANOVA). We offer a “word of caution” to avoid premature conclusions on group comparisons with statistical testing alone.

6. **Papers in this special issue**

This special issue draws on this renewed perspective on understanding individual differences, nuanced analysis and context affect our understanding of East Asian wisdom. We are proud to present the following contributions.

Michael Minkov presents “A revision of Hofstede’s model of national culture: Old evidence and new data from 56 countries”. Much of the measurement used to define Hofstede’s cultural dimensions have not changed since the original study in the 1970s: they are old and likely to be inapplicable today and in the future. Many measures could not be reproduced by more recent researchers. This study combines proprietary research with publicly-available data – the WVS 5 – to propose a shorter and more easily interpreted measurement system, with fewer dimensions. The paper argues that some of the original Hofstede dimensions such as uncertainty avoidance and the masculine/feminine aspect are “misleading artefacts”, and proposes an in-depth revision of Hofstede’s work, namely a reduction of dimensions.

Pauluzzo, Guarda, De Pretto and Fang, in their research paper, “Managing paradoxes, dilemmas, and change: a case study to apply the Yin Yang wisdom in Western organizational settings” offer a look at how cultural values and behaviour are managed in the organisations within Generali Group, which is one of the world’s largest insurance companies. Based on the East Asian wisdom of Yin Yang balancing and cultural perspective, this study explores how the West (i.e. Organisations and individuals) can be more effective in managing cultural differences by applying East Asia’s Yin Yang paradigm according to situation, context, location and time. The study demonstrates that integration and learning sourcing from opposite cultures leads organisations and individuals to manage and balance cultural paradoxes.
Hyun-Jung Lee and Carol Reade present “The role of Yin-Yang leadership and cosmopolitan followership in fostering employee commitment in China: a paradox perspective”. Lee and Reade’s paper looks at using a “paradox perspective”, also with a Yin Yang lens, with inspirations from for example, the earlier research on Yin Yang thinking and organisation paradox e.g., Peter Li (2016), Fang (2012), Faure and Fang (2008), Smith and Lewis (2011). This study explores leadership-followership dynamics in foreign firms in China. Yin Yang leadership behaviour (Japanese expatriates and Chinese employees) is tested for their roles in the formation of employee commitment. Yin Yang leadership and cosmopolitan followership “tango” together as cultural adaptability to build employee commitment, emphasising the interaction between leaders and followers.

Lin, P.P. Li and Roelfsema offer a valuable addition to the cross-cultural management literature with, “The traditional Chinese philosophies in inter-cultural leadership: The case of Chinese expatriate managers in the Dutch context”. The paper highlights important subtleties that Western authors have neglected. Specifically, using in-depth interviews, the authors test three propositions regarding the ease with which Chinese managers working in the Netherlands integrate with the local culture and workforce as a function of their influence by Confucianism, Taoism and Legalism. “Chinese expatriates’ leadership styles are simultaneously influenced by multiple traditional philosophies. This result is consistent with the theoretical argument that many Chinese, both historically and in modern times, tend to have a mix of multiple philosophical perspectives rather than being a believer in only a single philosophy”.

Much comparative management literature, naturally from the West, takes the perspective of a Western manager working in China, or some other Asian country. To see the perspectives of Chinese managers working in the West is refreshing.

Viengkham, Baumann and Winzar offer a first attempt at an empirical measurement of Confucian values applied to the work environment with “Confucianism: Measurement and Association with Workforce Performance”. The paper recognises that values are a relative construct, applying at different levels in different circumstances. Viengkham and her colleagues reconsider the measurement of Confucianism, or Confucian values, and subsequently probe associations with workforce performance. Different “shades of Confucianism” in East Asia are categorised into Confucian Origin, Preservation, and Pragmatism. That multifaceted measurement of Confucianism is next linked to workplace performance, demonstrating that indeed Confucianism drives human performance.

Xin Li’s paper, “Zhong-Yong as dynamic balancing between Yin-Yang opposites” expands on a recent commentary by Li et al. (2017) on Peter Li’s (2016) CCSM article titled “Global implications of the indigenous epistemological system from the East: How to apply Yin-Yang balancing to paradox management”. Zhong-Yong is the Chinese term to capture the “dynamic balancing between Yin Yang opposites”. Xin Li believes that Peter Li’s “asymmetry” and “superiority” arguments may be “flawed”, and instead offers an alternative understanding: Zhong-Yong is more sophisticated than the ratio-based one. Balance is more than defining a ratio of activities that deliver a simple average of competing perspectives: Balance is about selecting what is good, appropriate and feasible from competing options and discarding what is none of these, and then attempting to find a path to deliver as many as possible. Xin Li also draws a comparison between East and West, or how the “analysis plus synthesis” methodology would be like the Western “get the best of both worlds”.

The overarching themes that emerge in this Special Issue are Yin Yang, with the ideas of how different cultures deal with paradox, and Zhong-Yong, with accompanying concerns of how to conceptualise and deal with balance of opposites. Two themes implicit in many of these papers is how we might make empirical measures of our constructs, and how the interaction of culture and behaviour is moderated by context.
7. The ReVaMB model
We suggest an extension of Baumann and Winzar’s (2017) ReVaMB model from a Confucian orientation to a broader East Asian philosophical perspective. Culture, at least in the short term, does indeed determine, or “drive”, attitudes and behaviour, but this effect depends on circumstances, the environment; the broader context.

The model rests on three defensible assumptions:

1. Personal values and culture drive behaviour, at least in the very short term.
2. Different components of values and culture become activated, or take dominance in different circumstances. That is, context moderates the relationship between culture and behaviour.
3. Absolute measures of values or culture are not meaningful – it is the relative value of these constructs that affect decision-making and behaviour.

Our original formulation centred on Confucianism and its effects on workplace behaviour, specifically work ethic. There is no reason why this model cannot be generalised to include many philosophical, cultural or political values structures, and their effects on other specific behaviours. That is, East Asian wisdom such as Confucianism, Taoism, Buddhism and Legalism, determine behavioural outcomes. For example, Qiu and Rooney (2017) suggest a model of Buddhist mindfulness and workplace ethical behaviour can be operationalised very easily with this approach. We define moderating factors as the “context” for a situation to occur, or in statistical terminology, indeed moderators between independent variables such as East Asian wisdom (Confucianism, Taoism, Buddhism and Legalism) and the dependent behavioural or attitudinal variable. The model is visualised in Figure 7.

We invite future researchers to test, verify, extend etc. our original ReVaMB model on East Asian (and other) philosophical perspectives and context. What is the relative contribution of Confucianism, Taoism, Buddhism and Legalism (and other East Asian and other philosophies, ideologies and religions) to explain and predict behaviour and attitudes?

Source: Baumann and Winzar (2017)
References


Eichhorn, B.R. (2014), Common Method Variance Techniques, Cleveland State University, Department of Operations & Supply Chain Management: SAS Institute Inc., Cleveland, OH.


Further reading

Corresponding author
Chris Baumann can be contacted at: chris.baumann@mq.edu.au