



School of Management, Royal Holloway University of London

Working Paper Series

SoMWP-1004

**HOW NOT TO DO CROSS CULTURAL ANALYSIS: PREDICTIVE
FAILURE AND CONSTRUCTION FLAWS IN GEERT HOFSTEDE'S CASE
STUDY**

**Professor Brendan McSweeney, Dr. Donna Brown and
Dr. Stravroula Iliopoulou**

June 2010

The School of Management Working Paper Series is published to circulate the results of on-going research to a wider audience and to facilitate intellectual exchange and debate. The papers have been through a refereeing process and will subsequently be published in a revised form. Requested for permission to reproduce any article or part of Working paper should be sent to the publisher of this series.

The School of Management, Royal Holloway, University of London has over 65 academic staff who are organised into different research groups. The current research groups are:

Accounting, Finance and Economics

Strategy and International Business

Marketing

Technology and Information Management

Organisation Studies and Human Resource Management

The School also has around 60 research students attached to the various research groups. A strong theme of research in the School is its international and comparative focus.

Copyright © Professor Brendan McSweeney, Dr. Donna Brown and Dr. Stravroura Iliopoulou

ISBN: 978-1-905846-45-0

Published by:

The School of Management,
Royal Holloway University of London
Egham Hill
Egham
Surrey TW20 0EX

**HOW NOT TO DO CROSS CULTURAL ANALYSIS: PREDICTIVE
FAILURE AND CONSTRUCTION FLAWS IN GEERT HOFSTEDE'S CASE
STUDY**

*Professor Brendan McSweeney, Professor of Management, School of Management, Royal Holloway University of London, Egham, Surrey, TW20 OEX, United Kingdom.

Email: brendan.mcsweeney@rhul.ac.uk

Dr. Donna Brown, Senior Lecturer in Employment Relations & Managerial Economics, School of Management, Royal Holloway University of London.

Dr. Stravroula Iliopoulou, Lecturer in Finance, School of Management, Royal Holloway, University of London

* Corresponding author

ABSTRACT

This paper considers Hofstede's claim that his national cultural "dimension scores" and related rankings of countries have predictive power by examining a 'validating' case study he has used in a number of publications to 'demonstrate' that capability. When tested against cross-sectional and longitudinal empirical evidence the case study is shown to have neither explanatory nor predictive power. A further unpacking of the case study demonstrates methodological flaws in its construction. Some characteristics of valid cross-national case studies are then outlined in a discussion of these flaws.

Keywords: case studies; femininity; Hofstede; industrial relations; national culture; masculinity.

HOW NOT TO DO CROSS CULTURAL ANALYSIS: PREDICTIVE FAILURE AND CONSTRUCTION FLAWS IN GEERT HOFSTEDÉ'S CASE STUDY

INTRODUCTION

'National culture' is a highly contested notion. Its existence; its positing as a values or psychological core; the capacity of attitudinal surveys or other methods to measure or compare 'it' cross-nationally; and the appropriateness, adequacy, and accuracy of specific representations and rankings of national cultures, or of their differences; have all been extensively critiqued (Oyserman, Coon and Kimmelmeier, 2002; Kitayama, 2002; McSweeney, 2002; Tung, 2008, for instance). As Michael Hechter states: "our capacity to accurately assess these values is unimpressive" (2000: 697). But national culturalist assert that their favoured representations of national cultures (or national cultural differences) enable effective predictions of social action. Geert Hofstede, for instance, boldly claims that he has identified: "five main dimensions along which the dominant value systems in more than 50 countries can be ordered and that [they] *affect* human thinking, feeling, and acting, as well as organizations and institutions, in *predictable* ways" (Hofstede, 2001: xix)(emphasis added)(see also Hofstede and Hofstede, 2005: 31). Has Hofstede¹ provided "a system of generalisations that can be used to make correct predictions" (Friedman, 1953: 4)? If he has done so, his national cultural dimension indices, whether deemed accurate representations or not, would clearly be useful.

An evaluation of the predictive power of Hofstede's indices of national cultural differences is undertaken in this paper by testing Hofstede's claim that the "masculinity-femininity dimension affects ways of handling industrial conflicts" (Hofstede and Hofstede, 2005: 143)(see also Hofstede, 1991: 92; Hofstede, 2001: 316). Hofstede defines 'masculinity' versus 'femininity' as: "[A]ssertiveness and competitiveness versus modesty and caring" (Hofstede and Peterson, 2000: 401)(see also Hofstede, 2001: 297; Hofstede and Hofstede, 2005: 116). In the workplace in 'masculine' countries, he states, there is an emphasis on assertiveness, and in 'feminine' countries there is a preference for compromise and negotiations. His claim is first tested in this paper against a decade of data on "industrial conflict" (Hofstede and Hofstede, 2005: 143). It is then cross-sectionally and longitudinally tested against another indicator of the level of aggression within countries: homicide rates.

A VALIDATING CASE STUDY

Hofstede, in common with many employers of the notion of national culture, seeks, in part, to validate, his claim about the causal impact of the level of "masculinity-femininity" of a country with a number of case studies. The term 'case study' is used in many different ways (Ragin, 1992). In this paper it is used in the sense of a theoretically induced claim about how general social forces produce results in specific settings.

His seminal works (1980, 2001) are peppered with such cases. One he has reproduced in a number of his publications is as follows:

The masculinity-femininity dimension [of a national culture] affects ways of handling industrial conflicts. In the United States as well as in other masculine cultures (such as Britain and Ireland), there is a feeling that conflicts should be resolved by a good fight: "let the best man win." The industrial relations scene in these countries is marked by such fights. If possible, management tries to avoid having to deal with labor unions at all, and labor union behaviour justifies their aversion. In feminine cultures like the Netherlands, Sweden and Denmark, there is a preference for resolving conflicts through compromise and negotiations (Hofstede and Hofstede, 2005: 143) (see also Hofstede, 1991:92; 2001:316).

The case study asserts causal regularity: "[t]he masculinity-femininity dimension [of a national culture] affects ways of handling industrial conflicts." A causal claim (*Y* because of *X*) can be distinguished from a weaker notion of association or correlation: when *X* then *Y*. Association, as an immense classical and contemporary literature convincingly demonstrates, is not sufficient evidence of causality. But without association as a regular sequence causality cannot be validly said to exist nor can predictability be demonstrated. Regular association is thus a necessary but not sufficient condition of valid causal claims or effective predictions.² But as we shall see, there is not even a weak association between the supposed independent variable/cause (national gender as measured by Hofstede) and dependent variable/outcome (industrial conflict). The asserted associations/predictions, and thus the causal claim, in the case study are shown to consistently fail.

Following those tests, methodological flaws in the construction of Hofstede's case study are identified and examined, including: non-equivalence between the compared sets of countries; incompatibility with readily available data; assertion, not demonstration, of causality; and inconsistency within the causal claim. Drawing on those failings, some guidelines for more appropriate and valid cross-national comparative case studies are set-out.

PREDICTIVE POWER?

At what level should the predictive power of Hofstede's dimensions be tested? Many cross-cultural courses, training programmes, and multiple publications wrongly suppose that national averages (Hofstede's or others') also describe and can predict practice at levels 'lower' than the national – the organizational, the local, the individual, and so forth – where most practices occur. But making direct translations of properties or relations at one level to another, by projecting from a higher level to a lower (from the national to organizational or individual) – is unwarranted even if we suppose that the depiction of the national level is accurate. Robinson (1950) originally described the attribution of views about the characteristics of one level to other levels also as the "ecological fallacy" (1950) and Galtung calls it "the fallacy of the wrong level" (1967)(see also Hofstede, 2001: 16, 463). Relationships identified at one level of analysis may be stronger or weaker at a different level of analysis, or may even reverse direction (Klein and Kozlowski 2000). Disaggregation leads to

misrepresentation whenever populations are not wholly homogeneous. But error may also occur when a property at one level are attributed to a homogeneous group at a lower level. Schwartz (1994a), citing, Zito (1975), gives the illustrative example of the discrepancy between a hung jury at two levels. As a group, a hung jury is an indecisive jury, unable to decide the guilt or innocence of the accused. However, attributing that characteristic to the individual members of the jury would be incorrect as the jury is hung because its individual members are very decisive – not indecisive.

Hofstede states that the four dimensions he employed (masculinity-femininity, and so forth) “together account for 49% percent of the variance in *country mean scores*” of answers to an IBM employee survey – his primary data source (Hofstede, Neuijen, Ohayv, and Sanders, 1990: 288)(emphasis added). Even if we accept the accuracy of his calculations 51% of national variance in mean scores of respondents’ answers is unexplained. And of itself the 51% explains nothing about national behaviour. The analysis was of questionnaire answers – not of behaviour, actions, or practices. But even more significantly, in terms of level of analysis, the pattern of correlation found in national averages is *not* replicated at the individual level. Gerhart and Fang (2005: 977) estimate, based on Hofstede’s data, that only “somewhere between 2 and 4 percent” of the variance at the level of individuals answers is explained by national differences – a tiny portion. Hofstede’s own estimate of 4.2 per cent is only marginally higher (2001: 50). Furthermore, two of the four (later five) dimensions employed by Hofstede to depict national cultures – “power distance” and “individualism and collectivism” were statistically identified by him *only* in nationally averaged data. At the level of individuals they had near-zero intercorrelations (Bond 2002; Schwartz 1994b) for those dimensions and thus no explanatory power at that level. The massive gap between the ability of national-level data to describe or predict micro-level behaviour (above) is also consistent with the personality psychology literature which has long found that hypothesized global trait dispositions like friendliness, power-distance and dominance typically account for no more that 9 to 15 per cent of diversity of individual differences over naturally occurring situations (Shweder 1979).

In this paper we test the predictive power of some of Hofstede’s national cultural depictions at the level they should be most powerful: the macro-comparative (Bollen, Entwisle and Alderson, 1993) or national aggregate level. National level data smoothes-out local variations. More powerful, and more useful predictions, would be about subnational sites of action, for example, about actions within regions, sectors, or individual organizations. If reasonably accurate predictions could be made about conditions or behaviours at such sub-national levels that would indicate considerable national uniformity of practices and an immense predictive ability of Hofstede’s indices. However, in this paper the less demanding requirement – national level predictions only - is tested. A strong test and a weaker test of predictive power at that level are applied.

Strong test - comparative ranking: This test considers whether there is an association between a country’s ranking in Hofstede’s Masculinity Index (hereafter MAS Index) and the comparative level of industrial conflicts. The higher a country’s ranking (that is, the more ‘masculine’ it is rated on Hofstede’s MAS index) the greater should be the industrial conflict in that country. And conversely, the lower in the Index a country, the more ‘feminine’ it is deemed to be and the less intense should be such

conflict. Thus, of the six countries named in Hofstede's case study, Ireland (ranked joint 9th-10th in the MAS index) would have comparatively more disputes than Great Britain (ranked joint 12th with two other countries) which would have more disputes than the United States (ranked 19th). Similarly, the lower a country is rated in the MAS Index (that is, the more "feminine" a country) the fewer disputes it should have. Thus Sweden (ranked 74th) would have comparatively fewer disputes than the Netherlands (ranked 72nd), which would have fewer disputes than Denmark (ranked 71st).

The considerable variation over time in the levels of industrial disputes in countries might seem to readily falsify this deterministic notion of national culture, but of itself it does not. What is claimed/tested are not absolute levels of disputes but rankings: cross-national comparisons.

Weaker test - non-ranked dichotomy: This test considers whether there is a general association between a country depiction in the MAS Index as a 'masculine' or a 'feminine' country and the level of industrial conflict. A positive result would be that, whilst not necessarily in rank order, "masculine" countries would have more aggressive industrial relations than 'feminine' countries. So, for example, in the case of the six countries named by Hofstede, the requirement is merely that the three countries with the highest level of disputes are "masculine" and the three with the lowest level of industrial disputes are "feminine". It would not matter therefore, for instance, if Ireland, with the highest comparative 'masculinity ranking', had a lower level of disputes than one or both of the other two 'masculine' countries provided there were more disputes in Ireland than in the three 'feminine' countries.³

DATA

What data is appropriate for identifying the degree of conflict in industrial relations? Industrial relations are complex. But strikes and lockouts are widely regarded as a good measure of the level of industrial conflict in a country (provided there are not significant coercive restrictions on the right to strike)(Chernyshev, Egger, Mehran and Ritter, 2002). However, absolute measures are of little use for international comparisons because of the great differences in sizes of countries. There is a wide consensus, that the best available comparative indicator of levels of conflict in industrial relations is working days not worked due to labour disputes per thousand employees (Edwards, 1995).

Tables 1 and 2 (below) shows data of days lost due to labour disputes per 1,000 employees over a ten year period (1993 to 2005, inclusive). The data is divided into two five year periods (1996-2000, inclusive) and (2001-2005, inclusive) in all industries and services for the three 'masculine' countries and three 'feminine' countries named in Hofstede's case study, namely: 'masculine' Ireland, Britain, and the United States and 'feminine' Sweden, Netherlands, and Denmark.

The stronger, the comparative ranking, test is first discussed.

The first column on the left in Table 1 names the countries (from top to bottom of the table) in the ranking order predicted on the basis of the MAS Index. The most

masculine country in the Index, Ireland, is first, and so on down to the most feminine of the six countries, namely, Sweden. The two columns to the right show the actual rankings based on working days not worked due to labour disputes per 1,000 employees in all industries and sectors. As the most 'masculine' of the six countries, Ireland should have the highest level of disputes. But instead the country with the highest level of disputes in both of the periods is 'feminine' Denmark.⁴ Great Britain predicted to be the second most aggressive country was in fourth place in each of the periods. Netherlands predicted to be more aggressive than Sweden on the basis of the MAS Index, had comparatively fewer disputes than in Sweden in all three periods. Sweden the most 'feminine' of the six countries – indeed the most 'feminine' of all countries in Hofstede's MAS index had the second highest level of disputes in 2001-2005. Of the twelve rankings in Table 1 only one is predicted correctly, that of the United States in 1996-2000. In neither of the two periods does the actual ranking of the six countries named by Hofstede match the ranking predicted on the basis of his MAS index. A considerable failure of prediction. An analysis of data for each individual year within 1996-2005 (inclusive) also shows that in *none* of the years does Hofstede's ranking match the actual six-country ranking ([UK] Office of National Statistics, 2007; 2002).

Clearly the MAS Index fails the stronger test. But what of the weaker non-ranked dichotomy test? Does the mere classification of a country as 'masculine' or 'feminine' in the MAS Index have any predictive power? Ignoring ranking within masculine or feminine countries, is there more industrial conflict in 'masculine' countries than in 'feminine' countries?

In neither of the two periods are the three countries with the highest levels of industrial disputes all 'masculine' and in none of those periods are all of the countries with the lowest levels of industrial disputes 'feminine' (Table 1; Figure 1). A decisive failure. In the ten individual years (1996 onwards to 2005) in only one (1996) were the top three countries 'masculine' and the bottom three 'feminine' (albeit with a different ranking than predicted by Hofstede)(Office of National Statistics, 2007; 2002). So, in twelve out of thirteen years, Hofstede's case study even fails the weaker category test. A decisive failure.

Unrepresentative countries. The claim in Hofstede's case study is not merely that a causal relationship between national gender and industrial conflict exists in the six named countries, but in all countries. As we have seen - for the six named countries the causal claim is at odds with the actual record of industrial disputes. But even had the data for the six countries been in accord with Hofstede's causal claim that would not have been sufficient supportive evidence of an association between 'masculinity' and higher levels of industrial disputes. A necessary condition of valid comparison is that the comparators are equivalent. But the comparison in Hofstede's case study is not equivalent: 'feminine' countries are not compared with countries with equivalent levels of 'masculinity'. The named 'feminine' countries are at the extreme feminine end of the MAS Index: Sweden (most); Netherlands (3rd most); and Denmark (4th most). But these are not compared with any of the top four most 'masculine' countries in the Index. Indeed, only one of the masculine countries (Ireland) is in his list of the top ten most masculine countries. It's ranked joint 9th, whilst Great Britain and the United States are joint 12th and 19th respectively (Hofstede and Hofstede, 2005: 120-121).

To correct for that defect Table 3 (below) includes ten additional countries – five ‘masculine’ and five ‘feminine’. They were selected in accordance with the following criterion: the most ‘masculine’ and the most ‘feminine’ countries (if not already in Hofstede’s six named countries) for which reliable and comparable labour dispute data for the time periods under review was available. Unavailability of data excluded Slovakia which is the most ‘masculine’ country in the MAS Index. Instead, Japan ranked second highest for ‘masculinity’ in the MAS Index, and for which comparable data is obtainable, is included as is Norway ranked the second most ‘feminine’ country in Hofstede’s index. Also excluded is Hungary⁵ (third highest in Hofstede’s MAS index) as no data is available for the period 1996-2000. So, given the availability of data criterion, Tables 3 and 4 compare eight equivalent ‘masculine’ and eight ‘feminine’ countries.

Again, Hofstede’s predictions fail both the comparative ranking and the non-ranked dichotomy test. It is clear to the naked-eye that there is no association. Japan the most masculine country on the basis of the MAS index had the *lowest* number of disputes jointly with ‘masculine’ Austria and ‘masculine’ Luxembourg in 1996-2000; and the lowest jointly with Luxembourg in 2001-2005. In the ten-year period (1996-2005) no country had fewer disputes than Japan but the MAS Index based prediction is that it would have the highest. ‘Feminine’ Denmark predicted to be the 13th lowest had the highest level of disputes in 1996-2000 followed by ‘feminine’ Spain with the second highest level. In 2001-2005 ‘feminine’ Spain (predicted to be 10th lowest out of sixteen) had the highest level of disputes. Out of the 32 rankings in Table 3, based on actual levels of disputes, only 2 matched ranking predicted on the basis of the MAS index. In both periods five out of the eight highest countries for industrial disputes were ‘feminine’ (see Figure 2, below). The MAS index predicted outcome is that all eight would be ‘masculine’. Another failure of prediction.

POWER-DISTANCE

Hofstede also states that a country’s position in his Power-Distance Index (hereafter P-D Index)⁶ is a good predictor of the levels of violence in that country (2001: 111). “Smaller power distances”, he says, are associated with a certain consensus amongst the population that reduces disruptive conflicts” (2001: 111). Referring specifically to industrial relations he states that the P-D index “informs us about the dependence relationships in a country. In small-power-distance countries, there is limited dependence of subordinates on bosses, and therefore a preference for consultation” (Hofstede and Hofstede, 2005: 45). On this basis, the lower a country is in the P-D Index the comparatively more consensus (fewer disputes) is predicted for it. Table 5 (below) compares Hofstede’s predicted power-distance ranking of the same sixteen countries as in Tables 3 and 4 with actual ranking using the same industrial dispute data.

We can observe directly from Table 5 (below) that there is no association. For example, Denmark predicted on the basis of Hofstede’s P-D Index to have the second lowest record out of the sixteen countries, had the highest level of industrial disputes in 1996-2000 and the fifth highest in 2001-2005. Japan predicted to have the fourth highest level of disputes had the joint lowest in both periods. Austria predicted to have the lowest level of disputes did so jointly with two other countries in 1996-2000,

but it had the fourth highest in 2001-2005. Luxembourg predicted to be the sixth highest had jointly the lowest record in both periods. Out of 32 rankings based on actuals (Table 5) only 5 matched the predicted *outcomes*, yet another clear failure. The P-D Index has no predictive power in relation to “consensus” or aggression in terms of ranking of countries as measured by comparative levels of industrial disputes.

The P-D Index also fails the weaker non-ranked dichotomy test. In 1996-2000 eight of the sixteen countries and in 2001-2005 ten of the sixteen countries are miscategorised. For instance, in 2001-2005, Finland, Norway, Sweden, Ireland, Denmark and Austria predicted to be in the upper half in terms of higher levels of industrial disputes were in the lower half.

Correlations: Contrary to the claim in Hofstede’s case study the lack of predictive power of his MAS index in relation to industrial disputes is so great that it is clear to the naked-eye (Tables 1,2, 3, and 4, below) without the employment of more ‘sophisticated’ statistical analysis. Similarly, the explanatory and predictive ineffectiveness of his P-D Index in relation to industrial disputes is clearly revealed (Table 5, below). However, correlation tests were carried out on annual industrial dispute data for whole country data and on data sets for the production/construction sector, and that for the services sector.⁷ The data analysis revealed that there are no consistent statistically significant correlations between national industrial dispute data with the MAS Index or the P-D Index.

Interaction Term: Although Hofstede’s case study (above) refers only to one dimension (masculinity-femininity) theoretically it is possible within Hofstede’s model that strike rates could also simultaneously be affected by power-distance. To consider this an interaction term was created for the whole country sample. But no consistent significant correlations between this interaction term and successive industrial dispute rates were discovered.

ANOTHER INDICATOR OF AGGRESSION

The actual data on industrial disputes comprehensively refutes the causal claim in Hofstede’s case study above. As an additional test of Hofstede’s claim, that the comparative position of a country in terms of his measures of masculinity-femininity and/or power-distance indicates the comparative level of conflict in that country, these dimensions are looked at below in relation to another indicator of levels of conflict in a country: national rates of homicide. The proportions of a population who die in this manner differ dramatically between countries. Do Hofstede’s country rankings have predictive power in relation to homicide rates?

Table 6 (below) lists twenty-six countries ranked in order of annual average rates of homicides (excluding deaths due to capital punishment and operations of war) over a seven-year period 1990-1996 (inclusive). That ranking is compared with those predicted on the basis of the MAS and P-D indices. Countries included in the table are the twelve most ‘masculine’ and the twelve most ‘feminine’ countries in the MAS Index for which comparable homicide data is available together with any country named in Hofstede’s industrial relations validating case study (above) not included in

that list of twenty four. The only country thus initially excluded was the ‘masculine’ United States. It was included and matched with the nearest equivalent ‘feminine’ country for which comparable data was available, namely Spain. The United States is ranked as the nineteenth most masculine country in the MAS index, Spain is ranked as the twenty-third most ‘feminine’ country in the MAS with the score given to it by Hofstede close to that of the nineteenth most ‘feminine’ country in the MAS index: Croatia for which comparable homicide data was not available. In total there are thirteen matched ‘masculine’ and thirteen ‘feminine’ countries listed in Table 6 (below). The table is symmetrical in the sense that it contains the same number of equivalently ranked ‘masculine’ and ‘feminine’ countries.

The comparison shows considerable divergence between that the actual rankings for homicide and the rankings predicted using of either the MAS Index or the P-D Index. The country with the highest homicide rate - Columbia – is ‘masculine’. But, its predicted ranking using either of the indices is much lower - twelfth out of twenty-six countries on the basis of the ‘masculinity’ index and seventh on the basis of the power-distance index. The next two countries with the highest homicide rates (Russia and Estonia) are ‘feminine. Russia’s predicted P-D ranking is correct at second highest but its ranking on the basis of the MAS index is out by fourteen places. Japan which has the *lowest* homicide rate is predicted by its ‘masculinity’ ranking to be the country with the *highest* level of homicides and to be twelfth highest on the basis of its power-distance ranking. The top half of the table includes six ‘female’ countries out of thirteen. The bottom half of the table has six ‘masculine’ countries. The three countries with the lowest homicide rates, viz. Great Britain, Ireland and Japan, are ‘masculine’ – the exact opposite of the predictions based on the MAS Index.

The MAS Index also fails the weaker non-ranked dichotomy test. Instead of all masculine countries being in the most aggressive half (as measured by homicide rates) and feminine countries in the less aggressive half the actual results show mixed categories (Table 6; Figure 3). In the seven-year period covered in Table 6 almost half (six out of the thirteen) of the countries with the highest homicide rates were feminine.

Similarly, ranking in the P-D Index has no predictive power. Five of the countries in the top half of Table 6 were predicted on a power-distance basis to be in the bottom half of the table. Even for the remaining eight countries in the top half, overall there is no systematic relationship between their power-distance ranking and their rates of homicide. For instance, Slovakia which is predicted to have the highest homicide rate on the basis of both of Hofstede’s indices is only the eleventh highest whilst Columbia predicted to be seventh has the highest homicide rate. Of the fifty-two predicted rankings (Table 6), only three match the actual rankings. Another unambiguous failure.

The failure of Hofstede’s MAS index and his P-D index to be able to predict the homicide ranking of countries or even to be able to predict what is the ‘gender’ of countries with higher homicide rates and those with lower is, as was shown previously, evident from Table 6 and Figure 3. But correlation between these indices and homicide data were also calculated separately using both multi-year period data and single year data. The results were the same.

Notably, our results fail to signify any significant differences (in means and medians) between the male and female countries and between the low and high P-D ranked countries. If the MAS and P-D indices were meaningful and informative, we would expect that the groups (those higher or lower in the indices) would differ significantly in the rates of homicide documented for at least some of the examined periods of time. They do not.

Failure: The predictive power of two of Hofstede's national cultural rankings (masculinity-femininity and power-distance) was tested against data on actual events relating to industrial disputes and to homicides in multiple countries. In contrast with Hofstede's claims about the predictive capability of his national cultural rankings, none was found. There are no significant associations; positive correlation; no predictive power; and therefore no evidence of causality or otherwise useful information. These findings are consistent with geographer Philip Wagner wider view that: "Aggregating mightily, one can speak of national cultures. The chief attribute of such a broad concept is its uselessness (1975:11).

DISCUSSION AND CONCLUSIONS

As we have seen the flaws in Hofstede's industrial relations case study cannot be attributed to a data collection or analysis error – it suffers from much more severe methodological problems. The gap between actual data and the predicted outcome is so enormous that we can reasonably describe the case study as a projective fiction.⁸

Case studies play significant research and pedagogic roles not only in the management and business disciplines but also in the wider social sciences and the natural sciences (David, 2005). However, although case study research is widely undertaken, its ability to generate valid knowledge is often questioned. The central criticism is that case studies cannot spawn explanations or theory which is generalizable (Campbell and Stanley, 1981; King, Keohane and Sidney, 1994), that is, their findings cannot be shown to be universally valid. And yet, "much of what we know about the empirical world is drawn from such studies" (Gerring, 2004: 341). And arguably, such research is often at the forefront of theoretical development (Eisenhardt and Graebner, 2007; Ragin, 1992). The dismissal or downgrading of case study research because of its inability to generate generalizations has been widely rejected as an inappropriate, unnecessary, or unattainable criterion (see Geertz, 1970; Hamel, 1992; Kennedy, 1976; Mink, 1968; Murphy, 2003, for example). Case studies can achieve things which other research approaches cannot do, or do as well. But there are matters on which case studies have lesser, little, or no contribution to make. Just as there is no reason for large-scale variable analysts to apologise for the limitations of their methods so too the inability to generalize, in the sense of making universal statements, from a case study should not blind us to the contributions such studies can make. In any event, although general theory ("covering laws", in Carl Hempel's term) is sometimes seen as the ultimate standard for research, even theories deemed to have such a quality are usually insufficiently specific to guide policy or enable effective interventions at the organizational level. Generalizations often fade when we look at particular situations.

What lessons for cross-national case study research and writing can we learn from an analysis of Hofstede's industrial conflict case study? How in making cross-national comparisons can we avoid invalid inferences of the type made by Hofstede?

1. ***Ensure comparator equivalence***: A cross national case study is by its nature not a study of a single setting but of multiple settings – a minimum of two – and is therefore comparative. Validity requires that the cross-level comparators are equivalent - or to put it another way - that there is sampling frame comparability. As discussed above, in Hofstede's case study the three named feminine countries are compared with three masculine countries all of whom did not have an equivalent degree of masculinity in the MAS Index. In comparisons, including cross-national comparisons, sample composition is arguably more significant issue than sample size (Bollen, Entwisle and Alderson, 1993). Construct and measurement equivalence are also vital for valid cross national case studies (Mullen, 1995; Thomas, Hult, Ketchen Jr., Griffith, Finnegan, Gonzalez-Padron, Marmancioglu, Huang, Talay, and Cavusgil, 2008;) but these issues are not directly pertinent to Hofstede's case study.

2. ***Beware of confirmatory bias***: If the record of industrial relations was/is consistent with Hofstede's law-like generalisation, naming the six countries as *illustrative examples* would be perfectly appropriate as a means of communication. But they are named as supportive evidence. And yet as we have seen above the causal claim in Hofstede's case study is readily disconfirmed by actual data. Even in the countries named as examples in Hofstede's case study the record of industrial disputes (and homicide) is not consistent with Hofstede's assertion. There is thus evidence of confirmatory bias (positive test strategy)(Sloman, 2005; Klayman and Ha, 1987) – a disproportionate imposition of the researchers' prior beliefs. As King, Keohane, and Sidney (1994; 21) observe: "Any intelligent scholar can come up with a plausible theory for any set of data after the fact, yet to do so demonstrates nothing about the veracity of the theory."

Cross national researchers may be more familiar with one of the compared countries than with others. This may strengthen the presumption of a relationship between two factors and thus predisposes one to find and overweight evidence of that relationship and decrease the chances of finding disconfirming evidence and if found to discount it (Chapman and Chapman, 1969; Goldberg, 1968; Nisbett and Ross, 1980). For instance, studies suggest that many physicians are not good at revising their initial diagnosis to take account of later diagnostic tests (Berwick, Fineberg, and Weinstein, 1981); jurors often decide their verdict early in a trial process (Devine and Ostrom, 1985). We academics as people are not immune from these biases. In the aggregate, the evidence seems compelling that the human tendency is to look for evidence that is supportive of hypothesis we favour. It is, of course, not possible to examine a situation uninfluenced by categories, theories, hunches, and so on. Cases are made by invoking theory, implicitly or explicitly. But the results of case study research need not be overly predetermined by the prejudice we project onto the study of the case situation. We can and should test the results of our inevitable biases rather than allowing them to determine - unquestioned - our 'findings'. A research focus is not the same as a fixation on predetermined research findings.

Not all cross-national comparisons can be, or need be, empirically tested. Cross-national comparisons can serve a variety of useful purposes including: descriptions of differences or similarities; hypotheses or puzzle generation; challenging ethnocentrism; stimulate further studies; provision of new data; identification of the locational limits of existing theory. Academic research is not the only pathway to understanding. A good novel, a good play, even a good joke can help us understand, but they rest on different standards of evidence. Reader engagement and plausibility are desirable attributes of academic research, but they are poor guides to the validity of causal explanations. Research which makes causal claims about the empirical world requires more than a subjective sense of conviction. Empirical claims should have been empirically tested.

3. *Test against readily available information*: Every time a ‘magic word’ is muttered before table salt is immersed in warm water, the table salt will dissolve. But, of course, the salt will also dissolve without the ‘magic word’. The lesson is that correlation, or more mundanely the observance of coincidence in the same time or space does not validly show causality. In constructing a case study a writer should consider whether there are broadly similar circumstances in which the provisional belief about causality does not, or does, hold. Have similar circumstances had different and/or similar outcomes than in the case situation under investigation? Do similar outcomes to that of the case have different and/or similar circumstances?

We do not have to fully accept Karl Popper’s theory of falsification to be able to accept that data apparently inconsistent with a causal theory is a powerful challenge. Such data is not necessarily a decisive “refutation” (1979) requiring the abandonment of the theory, but it does demand a response and a modification or contingent justification of the initial explanation (Quine, 1953). Reassessment rather than rejection might be appropriate. If, however, a causal claim is posed in deterministic terms: a given set of conditions inevitably led to a specific outcome – a negative finding, provided one is confident in the accuracy of the data - is fatal to the causal claim (Dion, 1998).

The claims in Hofstede’s industrial relations case study are contradicted by *readily available* data. The data on industrial disputes (used in the tables above) is available without difficulty from a number of sources including: the International Labour Office; the Organization for Economic Cooperation and Development; and various offices of national statistics. But without even accessing that data, Hofstede’s causal claim is plainly contradicted by the well-known record of very low levels of industrial disputes in Japan and Germany. Throughout the post-2nd World War the industrial relations in those two countries has been the exemplar of co-operation (Thelen, 1991). And yet, in Hofstede’s ‘masculinity’ index, Japan is the second most masculine country and Germany the twelfth.

4. *Test historically*: Consistency of the degree of force is not a necessary characteristic of a causal force. But Hofstede’s, and many others’, notion of national culture is of a constant force: “[w]hile change sweeps the surface, the deeper layers remain stable, and the [national] culture rises from its ashes like the phoenix” (Hofstede and Hofstede, 2005: 36). “[N]ational values” are “given facts, as hard as country’s geographic position or its weather” (Hofstede and Hofstede, 2005: 13). The masculine-feminine differences he has set out in his index of countries “are”, he

states, “unlikely to disappear in the future” (1998: 4) - they are “a basic and enduring anthropological fact about a national society” (1998: 10) – “there is no sign of convergence of country cultures in the direction of more masculinity, nor in the direction of more femininity” (1998: 27).

For Hofstede his dimension measurements are “permanent causes” (Mills, 1843). But the record of industrial disputes (and much more besides) shows considerable variation over time. In 1996 the three ‘masculine’ countries named in Hofstede’s case study had comparatively more industrial disputes than the three ‘feminine’ countries, albeit not in the rank order predicted on the basis of the MAS or P-D indices. By supposing cultural constancy (and, if also acknowledged, that of other cultural and non-cultural influences) a generalisation can be made from data from just one-point-in-time. But in the decade examined in this paper, 1996 was the exception, in none of the other nine years did this dichotomised pattern occur. Absolute levels and comparative rankings change as does the record within countries and in a direction unrelated to ‘gender’. For instance, between 1996-2000 and 2001-2005 there was an 89% drop in disputes in ‘feminine’ Denmark and yet a 500% increase in ‘feminine’ Netherlands. In ‘masculine’ Ireland there was a reduction of 73% but a rise of 233% in ‘masculine’ Germany.

The absence of stability over time in the measurements employed (whether national averages or not) indicates a key defect of one-point-in-time measurements which characterise the great bulk of cross-‘cultural’ studies (Oyserman, Coon and Kimmelmeir, 2002; Schetuch, 1967;). The considerable variation also indicates that even if it is supposed that national culture exists and has an influence it is clearly not the only, or the dominant, influence.

For practical and institutional reasons, historical studies in organizations are often difficult (March and Sutton, 1997) but these constraints scarcely apply to macro-comparative studies for which data is often readily available.

5. *Avoid excessive conflation:* Research designs almost invariably face a choice between knowing more about less and knowing less about more (Gerring, 2004). Hofstede’s case study has few, if any, of the strengths of either a good variable-orientated approach or a good case study orientated approach - and has many of their weaknesses. Variable-orientated investigations are usually conservative by design, rarely assigning cause unambiguously to one variable. But Hofstede’s case study considers only one independent variable/cause and attributes deterministic power to it. Good variable orientated studies emphasise probabilistic outcomes and consideration of alternative explanations as rejection of possible explanations plays an essential role in choosing the preferred explanation. In contrast, Hofstede’s case study is absolute and no alternatives are considered. And on the other hand, the depth of data/familiarity with the diversity and richness of specific circumstances, with the process and dynamics of cause and effect, possible only through a case study approach is also absent.

Societal level models of all types, not just the cultural, often lack clarity about causality (Oyserman and Uskul, 2008). A ‘cause’ is described (well or badly) as is the outcome(s). But the causal process, the linkages between cause and outcome is too often not unfolded for the reader at least. Instead of descriptions of situated causal

mechanisms the mere fact that two conditions exist in the same case study time and space together with a general causal theory is treated as sufficient evidence that one caused the other. Hofstede's case study reduces immense multi-layered complexity within countries to a single level, mechanical, and 'anorexic' process. It's like what Chomsky, in another context, called "reducing physics to meter reading" (1968). Another way of depicting this issue is to be wary of conceptual over-stretch. There is an inverse relationship between the compoundness of a concept and the number of cases attributable to, or covered by, it (Mahoney, 2004; Sartori, 1970;). Sub-national analysis will often demonstrate the information poverty of national averages and reveal considerable within country heterogeneity (Smith, McSweeney and Fitzgerald, 2008).

To take the example of industrial relations – the object of explanation/prediction in Hofstede's case study. There is an immense and scholarly literature on industrial relations – including extensive discussions of the multiple influences on industrial disputes. The within country variations in the occurrences of industrial disputes are consistent with the effects of multiple, changing, and interacting influences. In Ireland, for instance, in 2006 days lost due to industrial disputes in the construction sector accounted for 65% of total days lost. But in 1997 only 0.04% of days lost were in that sector. In 1997 financial and other business sectors accounted for 32% of days lost but in 1999, 2000, and 2001 no days were lost in those sectors because of disputes. National level data obscures considerable within-country variation. This was illustrated above in relation to industrial disputes. It can also be seen in relation to homicide. Homicide rates vary not only between countries and over time, but also within them. They vary immensely across different locations, socio-economic, gender, and ethnic groups (Gaines and Kappeler, 1997). Within the U.S. for example, in 2003 the annual homicide rate per 100,000 of the population in 2003 in the states of Louisiana and Maryland was 13.0 and 9.5 but at the other end the rates were 1.2 and 1.3 in Maine and South Dakota respectively. In the period 1999-2001 the average homicides per 100,000 of a population was more than five times greater in Washington D. C. than in San Francisco. Nisbett and Cohen (1996) found that among white men, homicide in response to insults occurs at rates several times higher in the southern U. S. states than in the northern states (Akerlof and Kranton, 2005; Fiske, 2002;). Sub-state analysis of homicide (and multiple other practices) would demonstrate further spatial heterogeneity at sub-county or sub-city and so forth (see Law, Serre, Christakos, Leone, and Miller, 2004, for instance).

Explanations of changing levels of industrial disputes, homicides, or whatever require not merely multivariate approaches but multivariate ones that are combinatorial. As Ragin (1987: 27) observes: "rarely does an outcome of interest to social scientists have a single cause ... social causation [involves] different combinations of causal conditions [and] specific causes may have opposite effects depending on context". As an illustration even a *preliminary combinatorial analysis* of industrial relations in Ireland would need to consider multiple and interacting endogenous and exogenous circumstances and changes including: the strong sectoral distribution of trade union membership – some highly unionised others scarcely so; the common educational background of many of the employees and managers; the dominant position of one trade union in the unionised sectors; the radical series of successive national pacts between government, employers, and trade unions; the rivalries between craft unions wholly based in Ireland and those with continuing affiliations to largely UK based trade unions; the 'brain drain' and its later reversal; scale and type of foreign direct

investment, and so forth (Brown, 1981; D'Art and Turner, 2005; Geary and Roche, 2001; O'Mahoney and Delanty, 2001; Sweeney, 2008).

Social phenomena are complex not merely because they are almost always the outcome of multiple variables but also because those variables can combine in a variety of ways and at different times. The combinatorial, often complexly so, nature of social causation makes identification of causation highly challenging and far beyond the capability of unilevel analysis even when the latter is well executed.

6. Test whether the case study is internally consistent: Hofstede's industrial relations case study is not internally consistent - a 'masculine' *national* culture is said to generate/indicate "a feeling that conflicts should be resolved by a good fight" (Hofstede and Hofstede, 2005: 143). But then this is inconsistently supposed to affect only *part of* a national population viz. "labor". In 'masculine' countries "labor" is said to want a fight, but management in the same 'masculine' countries is said to evade a fight. A 'culture' that is said to influence a section only of a national population is not a 'national' culture.

Overall, what insights about industrial relations do Hofstede's case study and its claims about national culture provide? At best none even about national averages. Instead the case study misdirects. Misleading international business guidance provided is of the type: if you invest in 'masculine' countries your business will be characterised by frequent industrial disputes and when a dispute actually occurs the predominant reason is national culture which cannot be changed. However, if you invest in a 'feminine' country industrial relations in your business will be characterised by consensus. As the data in Tables 1,2,3,4, and Figures 1 and 2 clearly show, that guidance bears no relation with the historical reality.

The analysis in this paper of Hofstede's industrial relations case study does not of itself necessarily constitute a falsification of Hofstede's model of national culture. He also relies on other 'supporting' evidence. But the severe divergence from actual data and the methodological flaws shown to characterise the construction of his case study should encourage caution about accepting his attempts to validate his theory with case-studies.

REFERENCES

- Adamopoulos, J. (2008) On the entanglement of culture and individual behaviour. In F. J. R. van de Vijver, D. A. van Hemert, and Y. H. Poortinga (Eds.). *op. cit.*: 27-62.
- Akerlof, G. A. and Kranton, R. E. 2005. Economics and identity in G. A. Akerlof (and co-authors) *Explorations in pragmatic economics*. Oxford: Oxford University Press.
- Berwick, D. M., Finberg, H. V. and Weinstein, M. C. 1981. When doctors meet numbers, *American Journal of Medicine*, 71(6): 991-8.
- Bollen K. A., Entwisle, B. and Alderson, A. S. 1993. Macrocomparative research methods, *Annual Review of Sociology*, 19(1): 321-351.
- Bond, M. H. 2002. Reclaiming the individual from Hofstede's ecological analysis – a 20-year odyssey: comment on Oyserman *et al.* *Psychological Bulletin* 128(1): 73-77.
- Brown, T. 1981. *Ireland: a social and cultural history 1922-79*, Glasgow: Fontana.
- Campbell, D. T. and Stanley, J. C. 1981. *Experimental and quasi-experimental designs for research*. Boston MA: Houghton Mifflin.
- Chapman, L. J., and Chapman. J. P. 1969. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74: 271-280.
- Chernyshev, I., Egger, P. Mehran, F. and Ritter, J. 2002. *Measuring decent work with statistical indicators*, Policy Integration, Statistical Development and Analysis, Working Paper No. 2, Geneva: International Labour Office.
- Chomsky, N. 1968. *Language and mind*. New York: Harcourt Brace Jovanovich.
- David, M. 2005. *Case study research, Vols. I, II, III, IV*. London: Sage.
- D'Art, D. and Turner, T., 2005. Union recognition and partnership at work: A New Legitimacy for Irish Trade Unions? *Industrial Relations Journal*, 36(2): 121-139.
- Devine, P. G. and Ostrom T. M. 1985. Cognitive mediation of inconsistency discounting. *Journal of Personality and Social Psychology*, 49(1): 5-21.
- Dion, D. 1998. Evidence and inference in the comparative case study. *Comparative Politics*. 30: 127-145.
- Eisenhardt, K. M. and Graebner, M. E. (2007) Theory building from cases: Opportunities and challenges, *Academy of Management Journal*, 50.1, 25-32.
- Edwards, P. 1995. Strikes and industrial conflict. In Edwards, P. (Ed.), *Industrial relations: theory and practice*, Oxford: Blackwell, 434-60.

- Feyerabend, P. 1975. *Against method*. London: New Left Books.
- Fiske, A. P. 2002. Using individualism and collectivism to compare cultures – a critique of the validity and measurement of the constructs: comments on Oyserman *et al.* *Psychological Bulletin*. 128.1: 78-88.
- Friedman, M. 1953. The Methodology of positive economic. In *Essays on Positive Economics*, Chicago: University of Chicago Press.
- Galtung, J. 1967. *Theory and methods of social research* Oslo: Univsitetsforlaget.
- Gaines, L.K. and Kappeler, V. 2003. *Policing in America*, (4th. Edition). Cincinnati: Anderson Publishing Co.
- Geary, J. F. and Roche, W. K. 2001. Multinationals and human resources practices in Ireland: a rejection of the 'new conformance thesis'. *International Journal of Human Resource Management*, 12(1): 109-107.
- Geddes, B. 1991. How the cases you chose affect the answers you get. *Political Analysis*, 2: 131-50.
- Gerring, J. 2004. What is a case study and what is it good for? *American Political Science Review*, 98(2): 341-354.
- Geertz, C. 1970. The impact of the concept of culture on the concept of man. In E.A. Hammel and W. S. Simmons eds. *Man makes sense*. Boston: Little Brown. 46-65.
- Gefland, M. J., Erez, M. and Aycan, Z. 2007. Cross-cultural organizational behavior. *Annual Review of Psychology*. 58: 479-514.
- Gerhart, B. and Fang, M. 2005. National culture and human resource management: assumptions and evidence. *The International Journal of Human Resource Management* 16(6): 971-986.
- Gilligan, J. 1996. *Violence: our deadly epidemic and its causes*, New York: Grosset/Putman.
- Goldberg, L. R. 1968. Simple models or simple processes? Some research in clinical judgment. *American Psychologist*, 23: 483-496.
- Hamel, J. 1992. On the status of singularity in sociology. *Current Sociology* 40(1), 99-119.
- Hechter, M. 2000. Agenda for Sociology at the Start of the Twenty-first Century, *Annual Review of Sociology*, 26: 697-98.
- Hofstede, G. 1980. *Culture's consequences: international differences in work-related values*. Beverly Hills, CA: Sage.

- Hofstede, G. 1991. *Cultures and organizations: software of the mind*. London: McGraw-Hill, 1991.
- Hofstede, G. 2001. *Culture's consequences: comparing values, behaviours, institutions and organizations across nations*. 2nd ed. Thousand Oaks CA: Sage.
- Hofstede, G. 2002. Dimensions do not exist – a reply to Brendan McSweeney. *Human Relations*, 55(11): 1355-61.
- Hofstede, G. 2006. What Did the GLOBE Really Measure? Researchers' Minds Versus Respondents' Minds, *Journal of International Business Studies*, 37: 882-896.
- Hofstede, G. and Associates (1998) *Masculinity and Femininity: The Taboo Dimension of National Culture*, London: Sage.
- Hofstede, G. and Hofstede, G. J. 2005. *Cultures and Organizations: Software of the Mind*. 2nd ed. New York: McGraw-Hill.
- Hofstede, G., Neuijen, B., Ohayv, D. D., and Sanders, G. 1990. Measuring organizational cultures: A qualitative and quantitative study across twenty cases. *Administrative Science Quarterly* 35: 286-316.
- Hofstede, G. and Peterson, M. F. 2000. National values and organizational practices, in N. M. Ashkanasy, C. P. M. Wilderom and M. F. Peterson (eds.) *Handbook of organizational culture and climate*, London: Sage, 401-405.
- Hollis, M. 1994. *The Philosophy of Social Science: An Introduction*. Cambridge: Cambridge University Press.
- Kennedy, M. M. 1976. Generalizing from single case studies. *Evaluation Quarterly*, 3: 661-678.
- King, G., Keohane, R. O. and Sidney, V. 1994. *Designing social inquiry: scientific inference in qualitative research*, Princeton University Press.
- Kitayama, S. 2002. Cultural psychology of the self: a renewed look at independence and interdependence. In C. von Hofsten and L. Bäckman (Eds.), *Psychology at the turn of the millennium, Vol. 2: Social developments and clinical perspectives*, New York: Taylor & Francis.
- Klayman, J. and Ha, Y-W. 1987. Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, 94: 211-228.
- Klein, K. J. and Kozlowski, S. W. (2000) From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods* 3: 211-236.

- Law, D. C. G., Serre, M. L., Christakos, G., Leone, P. A. and Miller, W. C. 2004. Spatial analysis and mapping of sexually transmitted diseases to optimize intervention and prevention strategies, *Sexually Transmitted Infections*, 80: 294-299.
- Liebersohn, S. 1985. *Making it count: the improvement of social research and theory*. Berkeley: University of California Press.
- Liebersohn, S. 1991. Small n's and big conclusions: an examination of the reasoning in comparative studies based on small numbers. *Social Forces*, 70: 307-20.
- MacIntyre, A. 1985. The character of generalizations in social science and their lack of predictive power, in A. MacIntyre, *After Virtue*: 88-108, London: Duckworth.
- McSweeney, B. 2002. Hofstede's model of national cultural differences and their consequences: a triumph of faith – a failure of analysis. *Human Relations*, 55(11): 89-117.
- McSweeney, B. 2009. Dynamic diversity: variety and variation within countries. *Organization Studies*, 30.9.
- Mahoney, J. 2004. Comparative-historical methodology. *Annual Review of Sociology*, 30: 81-101.
- March, J. G. and Sutton, R. I. 1997. Organizational performance as a dependent variable. *Organization Science*, 8(6): 698-706.
- Miller J. G. 2002. Bringing culture to basic psychological theory – beyond individualism and collectivism: comment on Oyserman *et al.* *Psychological Bulletin* 128: 97-109.
- Mink, L. O. 1968. Philosophical analysis and historical understandings. *Review of Metaphysics*, 20: 667-98.
- Miron, J. A. 2001. Violence, guns, and drugs: a cross-country analysis. *Journal of Law and Economics*, XLIV, 2 (pt. 2): 615-634.
- Mullen, M. R. 1995. Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 26.
- Murphy, K. R. 2003. *Validity generalization: a critical review*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Nisbett, R. E. and Cohen, D. 1996. *Culture of honor*. Boulder, CO: Westview Press.
- Nisbett, R. E. and Ross L. 1980. *Human inference: strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice-Hall.
- Office of National Statistics 2002. *Labour market trends*. London: The Stationary Office.

- Office of National Statistics 2007. *Economic and labour market review*. April, London: The Stationary Office.
- O'Mahoney, P. and Delanty, G. 2001 *Rethinking Irish history: nationalism, identity and ideology*. Basingstoke: Plagrave.
- Oyserman, D., Coon, H. M. and Kemmelmeier, M. 2002. Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*. 128(1): 3-72.
- Oyserman, D. and Uskul, A. K. 2008. Individualism and collectivism: societal-level process with implications for individual-level and society-level outcomes. In F. J. R. van de Vijver, D. A. van Hemert, and Y. H. Poortinga (Eds.): 145-173, *op cit*.
- Popper, K. R. 1979. *Objective knowledge: An evolutionary approach (Revised Edition)*. Oxford: Clarendon Press.
- Parboteeah, K. P., Hoegl, M. and Cullen, J. B. 2008. Managers' gender role attitudes: a country institutional profile approach. *Journal of International Business Studies*, 39(5): 795-813.
- Quine, W. V. 1953. Two dogmas of empiricism. In *From a logical point of view*. Harvard: Harvard University Press.
- Ragin, C. C. 1987. *The comparative method: moving beyond qualitative and quantitative strategies*. Berkley: University of California Press.
- Ragin, C. C. 1992. Introduction: cases of "what is a case", in C. C. Ragin and H. S. Becker (Eds.), *op.cit.*: 1-17. Cambridge: Cambridge University Press.
- Ragin, C. C. and Becker, H. S. (Eds.).1992. *What is a case? Exploring the foundations of social inquir.* Cambridge: Cambridge University Press.
- Robinson, W. S. 1950. Ecological correlations and the behaviour of individuals. *American Sociological Review*, 15: 351-357.
- Sartori, G. 1970. Concept misformation in comparative politics. *American Political Science Review*, 64: 1033-46.
- Scheuch, E. K. 1967. Society as context in cross-cultural comparisons. *Social Science Information*, 6(5): 7-23.
- Schwartz, S. H. 1994a. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *American Journal of Public Health*, 84: 819-824.
- Schwartz, S. H. 1994b. Beyond individualism/collectivism; new cultural dimensions of values. In U. Kim et al. (Eds.) *op. cit.*: 85-119.
- Shweder, R. A. 1979. Rethinking culture and personality theory part I: a critical examination of two classical postulates. *Ethos* 7(3): 255-278.

- Sloman, S. 2005. *Causal models: how people think about the world and its alternatives*, Oxford: Oxford University Press.
- Smith, C., McSweeney, B., and Fitzgerald, R. (Eds.). 2008. *Re-making management: between global and local*. Cambridge: Cambridge University Press, 2007.
- Sweeney, P. 2008. *Ireland's economic success: reasons and lessons*,. Edinburgh: Dunedin Academic Press.
- Thelen, K.A. 1991. *Union of parts: labour politics in postwar Germany*. London:Ithaca.
- Thomas, G., Hult, M., Ketchen Jr., D. J., Griffith, D. A., Finnegan, C. A., Gonzalez-Padron, T., Marmancioglu, N. Huang, Y., Talay, M. B. and Cavusgil, S. T. 2008. Data equivalence in cross-cultural international business research: assessments and guidelines. *Journal of International Business Studies*, 39(6): 1027-1044.
- Tung, R. L. 2008. The cross-cultural research imperative: the need to balance cross-national and intra-national diversity. *Journal of International Business Studies* 39(1): 41-46.
- van de Vijver, F. J. R., van Hemert, D. A. and Poortinga, Y. H. (Eds.). 2008. *Multilevel analysis of individuals and cultures*. New York: Lawrence Erlbaum Associates.
- Zito, G. 1975. *Methodology and meaning; variety of sociological inquiry*. New York: Praeger.

Table 1: Comparison of predicted rankings on the basis of the MAS Index with actual rankings

Country ranking based on annual averages of working days not worked due to labour disputes, per 1,000 employees in all industries and services

Predicted Ranking	1996-2000	2001-2005
1. M-Ireland	2	3
2. M-Great Britain	4	4
3. M-United States	3	5
4. F-Denmark	1	1
5. F-Netherlands	6	6
6. F-Sweden	5	2

Notes: M ('Masculine' country); F ('Feminine' country).

Sources: [UK] Office of National Statistics (2007); Hofstede and Hofstede (2005).

Table 2: Working days not worked due to labour disputes, per 1,000 employees in all industries and services

Annual Averages

Actual Country Ranking for 1996-2005	1996-2000	2001-2005
1. F-Denmark (4)	296	36
2. M-Ireland (1)	91	30
3. M-United States (3)	61	13
4. M-Great Britain (2)	21	26
5. F-Sweden (6)	9	34
6. F-Netherlands (5)	4	12

Notes: The number in a bracket after each country's name indicates its comparative ranking in the MAS Index.

M ('Masculine' country); F ('Feminine' country).

Sources: Office of National Statistics (2007); Hofstede and Hofstede (2005).

Table 3: Comparison of predicted ranking of sixteen countries on the basis of the MAS Index with actual comparative ranking

Predicted ranking	<i>Actual Ranking based on annual averages (%) of working days not worked due to labour disputes, per 1,000 employees in all industries and services</i>	
	1996-2000	2001-2005
<i>'Masculine' countries</i>		
1. Japan	14/15/16	15/16
2. Austria	14/15/16	4
3. Italy	5	2
4. Ireland	4	8
5. Great Britain	9	10
6. Germany	13	14
7. United States	7	13
8. Luxembourg	14/15/16	15/16
<i>'Feminine' countries</i>		
9. France	6	6/7
10. Spain	2	1
11. Portugal	10	11
12. Finland	8	3
13. Denmark	1	5
14. Netherlands	12	12
15. Norway	3	9
16. Sweden	11	6/7

Sources: Office of National Statistics (2007); Hofstede and Hofstede (2005).

Table 4: Working days not worked due to labour disputes, per 1,000 employees in all industries and services

<i>Annual Averages</i>		
Actual ranking based on annual averages in 1996-2005	1996-2000	2001-2005
1. F-Spain (10)	182	189
2. F-Denmark (13)	296	36
3. M-Italy (3)	76	120
4. F-Norway (15)	134	29
5. F-Finland (12)	56	91
6. M-Ireland (4)	91	30
7. F-France (9)	66	34
8. M-Austria (2)	1	80
9. M-United States (7)	61	13
10. M-Great Britain (5)	21	26
11. F-Sweden (16)	9	34
12. F-Portugal (11)	20	15
13. F-Netherlands (14)	4	12
14. M-Germany (6)	2	4
15. M-Luxemburg (8)	1	0
16. M-Japan (1)	1	0

Notes: The number in a bracket after each country's name is the predicted outcome on the basis of the MAS Index.

M ('Masculine' country); F ('Feminine' country).

Sources: Office of National Statistics (2004; 2007)

Table 5: Predicted ranking based on comparative position in the P-D Index compared with actual comparative ranking

*Country ranking based on Annual Average (%)
working days not worked due to labour disputes,
per 1,000 employees in all industries and services*

Predicted ranking	1996-2000	2001-2005
1. France	6	6/7
2. Portugal	10	11
3. Spain	3	1
4. Japan	14/15/16	15/16
5. Italy	5	2
6. Luxembourg	14/15/16	15/16
7. United States	7	13
8. Netherlands	12	12
9. Germany	13	14
10. Great Britain	9	10
11. Finland	8	3
12. Norway	2	9
13. Sweden	11	6/7
14. Ireland	4	8
15. Denmark	1	5
16. Austria	14/15/16	4

Sources: Office of National Statistics (2004; 2007); Hofstede and Hofstede (2005)

Table 6: Homicide rates per 100,000 of a population

<i>Actual ranking based on 1990-1996 homicide data</i>	MAS ranking	P-D ranking	1990-1996 (annual averages)	1993-1996
1.M-Columbia	12	7	84.18	83.50
2.F-Russia	16	2	24.60	29.98
3.F-Estonia	18	15	19.64	24.03
4.M-Mexico	7	3	17.50	17.27
5.M-Venezuela	5	4	14.60	15.30
6.M-United States	13	16	9.65	9.30
7.F-Costa Rica	21	18-20	5.18	5.55
8.M-Hungary	3	14	3.70	3.70
9.F-Finland	20	21	3.18	3.13
10. F-Chile	19	8	3.00	2.75
11. M-Slovakia	1	1	2.40	2.40
12. M-Italy	6	13	2.35	1.70
13. F-Slovenia	22	6	2.13	1.98
14. F-Portugal	17	9	1.54	1.50
15. F-South Korea	15	10	1.48	1.60
16. M-China	9	5	1.37	1.13
17. M-Austria	4	26	1.32	1.17
18. F-Sweden	26	23	1.24	1.18
19. F-Denmark	23	25	1.21	1.21
20. F-Netherlands	24	17	1.17	1.20
21. M-Germany	10-11	18-20	1.13	1.15
22. F-Norway	25	22	1.10	0.93
23. F-Spain	14	11	0.93	0.93
24. M-Great Britain	10-11	18-20	0.83	1.00
25. M-Ireland	8	24	0.66	0.65
26. M-Japan	2	12	0.60	0.60

Note: M ('Masculine' country); F ('Feminine' country).

Sources: Miron (2001); Hofstede and Hofstede (2005)

Figure 1: Predicted gender ranking category based on the MAS Index of the six countries compared with actual ranking (industrial disputes)

Predicted outcome	<i>Upper Half</i>	<i>Lower Half</i>
	MMM	FFF
Actual 1996-2000	FMM	MFF
Actual 2001-2005	FFM	MFF

Note: M: Masculine; F: Feminine.

Sources: Office of National Statistics (2007, 2002)

Figure 2: Predicted gender ranking category based on the MAS Index of the sixteen countries compared with actual ranking (industrial disputes)

Predicted gender	<i>Upper Half</i>	<i>Lower Half</i>
	MMMMMMMM	FFFFFFF
Actual 1996-2000	FFFMMFMF	MFFFMMMM
Actual 2001-2005	FMFMFFFM	FMFFFMMM

Note: M: Masculine; F: Feminine.

Sources: Office of National Statistics (2007, 2002); Hofstede and Hofstede (2005)

Figure 3: Predicted categorization based on the MAS Index compared categorization based on actual events (homicide rates per 100,000 of a population)

Predicted gender	<i>Upper Half</i>	<i>Lower Half</i>
	MMMMMMMMMMMMMM	FFFFFFFFFFFFFF
Actual 1990-1996	MFFMMMFMFFMMF	FFMMFFFMFFMMM

Note: M: Masculine; F: Feminine.

Sources: Miron (2001); Hofstede and Hofstede (2005)

¹ Hofstede national cultural research is one of the most cited in the Social Science Citation Index (Parboteeah, Hoegel and Cullen, 2008). It has become an almost emblematic citation in a number management disciplines.

² A classic example is in Yule and Kendal (1950: 315-316) who observed that there was a very high correlation (a correlation coefficient of 0.998) between the number of wireless receiving licences taken out from 1924 and 1937 in the United Kingdom and the number of notified mental illnesses for the same period (in Lieberman, 1985: 9).

³ Even the weaker test, if positive, would arguably not indicate that Hofstede's MAS index can provide some useful information. As discussed later in the paper, to have such positive content the two compared groups (masculine and feminine countries) need be equivalent – but those in Hofstede's case study are not.

⁴ Whilst there is considerable similarity between the coverage and methodology for data gathering in Denmark and Ireland, the minimum criteria for inclusion of a dispute in Denmark is more conservative than in Ireland and thus on an identical comparative basis the comparative industrial disputes position of Denmark over Ireland would be even wider (Office of National Statistics, 2007).

⁵ In the period for which data is available for Hungary (2001-2005) the average days lost in 'masculine' Hungary was *lower* than in any of Hofstede's six named countries, including the Sweden the most feminine country in the MAS index.

⁶ Power-Distance is defined as 'the extent to which the less powerful members of organizations and institutions within a country expect and accept that power is distributed unequally (Hofstede and Hofstede, 2005: 46).

⁷ The results are available upon request from the corresponding author.

⁸ The lack of an evidence-base also characterises a number of his other validating case studies. For instance, Hofstede states that: “[f]eminine countries believe in modest leaders” (2001: 388). Relying on anecdotes one could perhaps identify ‘feminine’ countries with “modest leaders”, but there is no systematic relationship. Almost half of the countries (Hofstede, 2001) he deems to be ‘feminine’ have been controlled – in some cases for very prolonged periods – by dictators or highly autocratic leaders. For instance the following nine Hofstede deemed ‘feminine’ countries were controlled for lengthy periods by autocrats - Chile (Pinochet); Portugal (Salazar), Iran (Khomeini); Panama (Noriega); Romania (Ceausescu); Russia (Stalin), Serbia (Milosevic); Spain (Franco), Taiwan (Chiang Kai-shek).