## Comparability in Meaning Cross-Cultural Comparisons
### Andrey Pavlov

## Introduction

The measurement of abstract concepts, such as personal efficacy and privacy, in a cross-cultural context poses problems of comparability in meaning, i.e. that the same concept could be understood differently by individuals with different cultural background. While this is not a new problem, as comparative quantitative research attests, a novel solution has been recently suggested by a group of researchers working from Harvard. Led by Gary King, a political scientist, the group developed a modeling technique to be used in parametric data in order to correct for the problem of comparability. Respondents are asked to assess their personal stance with regard to the concept at hand, and compare this assessment to a hypothetical situation involving the same concept as depicted in a series of vignettes. Unlike self-assessment, the vignettes provide invariant tool against which to rank the respondent's self assessment and at the same time provides us with means of comparing the rankings among respondents in the sample. Depending on the results, it is then possible to correct for incomparability in meaning. The paper will explore the use of this technique in our GPD International Survey. Two sets of vignettes were used, one deals with individual control of personal information, and the second with the extent to which privacy is respected in international travel.

## Why do we need to correct responses?

To make it easier for the reader to understand the problem, it will be explained in the framework of one of the questions of our survey: *Q2 "To what extent do you have a say in what happens to your personal information?"* There are four possible responses to this question (*"Refused"* and *"Don't know"* responses are dropped from analysis): *"Complete say"*, *"A lot of say"*, *"Some say,"* and *"No say"*. Here are the results of the survey for Q2:

Figure 1: Raw response frequencies

| | Country | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Brazil** | **Canada** | **China** | **France** | **Hungary** | **Mexico** | **Spain** | **USA** |
| **No say** | 27.3% | 17.0% | 10.3% | 9.5% | 31.0% | 9.3% | 15.3% | 16.7% |
| **Some say** | 37.2% | 51.0% | 17.3% | 30.3% | 41.4% | 50.4% | 52.0% | 53.7% |
| **A lot of say** | 18.2% | 22.0% | 50.5% | 24.5% | 13.4% | 21.9% | 13.4% | 17.5% |
| **Complete say** | 17.3% | 9.9% | 22.0% | 35.7% | 14.2% | 18.4% | 19.3% | 12.1% |

This table shows the frequencies of responses falling in each category by country. We see for instance that France has an high percentage of "Complete say", China has "A lot of say" frequency at least twice as high as it is in other countries, while Hungarians reported mostly "No say" and "Some say". Does this mean that, compared to other countries, French and Chinese people have much more say in what happens to their

personal information, and Hungarian people have little say? We should be very careful in interpreting these figures, since what the table actually shows is people's opinion, based on their intuition, rather than verifiable facts.

It is clear that the concept being measured, as well as the suggested categorical scale, is abstract and can be understood by people very differently, depending on their cultural background and perhaps other factors, such as age or gender. This creates inconsistencies in responses, and we can't compare them directly.

## Outline of the model

The statistical model suggested by G. King et.al. is an attempt to measure this incomparability. To do this, respondents are asked a series of vignettes. In the vignettes they answer the same question but with regard to other hypothetical people, presented in short stories. They give their assessment on the same categorical scale. Using the answers to the vignette questions, the model corrects the answer for self assessment, and here is how.

Figure 2: Statistical model
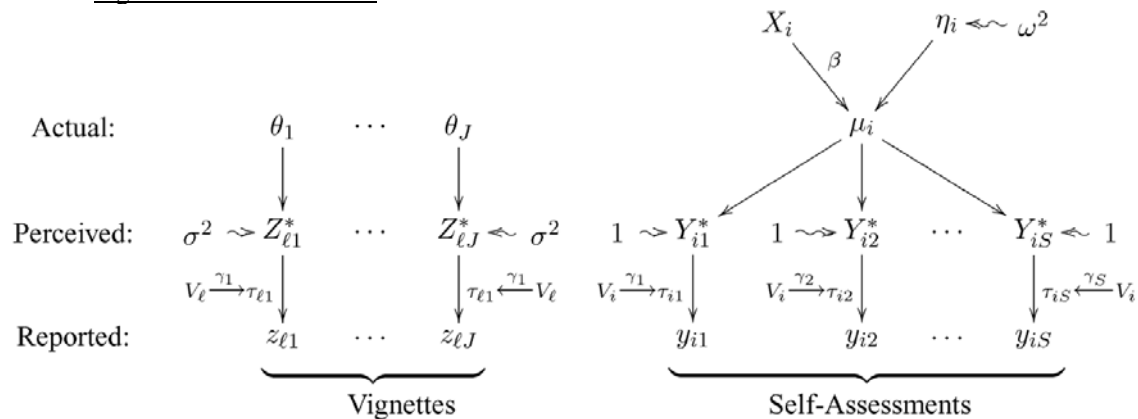
It is assumed by the model that there actually exists a variable (call it $\boldsymbol{\mu}$), measured on a continuous scale, which represents one's "amount of say" that he/she has in what happens to his/her personal information. We don't observe this variable of course, but we can perceive it and categorize it into one of the four responses. The perception is modeled by a normal error and categorization is done by means of three thresholds ($\boldsymbol{\tau}$'s), which divide the real numbers into four segments, corresponding to the four answers. Both the actual amount of say $\boldsymbol{\mu}$ and the thresholds depend on one's cultural background.
The vignettes, thought of as people, have their own "true" amount of say ($\boldsymbol{\theta}$'s in the diagram). When these $\boldsymbol{\theta}$'s are assessed by a respondent, we can estimate his/her thresholds, and then use them to estimate the respondent's $\boldsymbol{\mu}$.

## Preliminary analysis

Before we present the model for Q2, let's once again look at raw responses, summarized in table 1. To make the comparison of countries easier and following the ideas of the model, we develop a score which would represent how much say an average citizen of each country has in what happens to his/her personal information.

As in the model, assume that the "amount of say" is measured on a continuous scale, and that the four responses are equidistant on that scale. Assign values for the responses in the following way: "No say" = 0, "Some say" = 1, "A lot of say" = 2 and "Complete say" = 3. Now compute the mean of these values for each country. If we want to classify the resulting means into one of the four categories, we may take the midpoints between our codes as thresholds, that is 0.5, 1.5 and 2.5. Here is what we get:

Figure 3: What we might get without correcting incomparability
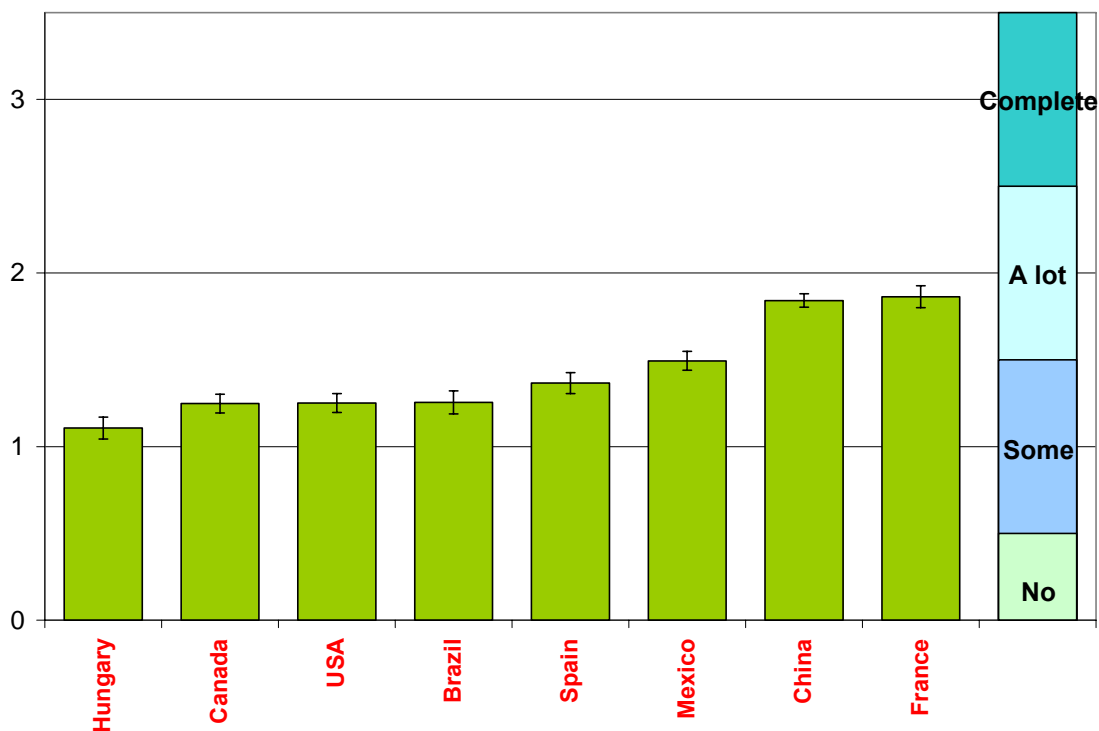


The error bars in the chart represent 95% confidence intervals for the estimated scores. As we have already observed, France and China received the highest score, while Hungary has the lowest score. Also if we look at our conventional categorical scale, we see that all countries are at least in the "Some say" region, with China and France scoring high in the "A lot of say" region, followed by Mexico, which is on the boundary. It is worth noting at this point, to be discussed further, that the response values were chosen somewhat arbitrarily, and so the heights of the bars are not very informative. Rather we should look at the differences, for instance we can note the big gap between China and Hungary or conclude that Canadian score is not significantly different from that of Brazil.

Now take a look at the vignette responses. We don't use the model at this point, just look at how people have rated the four vignettes.

Figure 4: Vignettes response frequencies

| Mike | Country | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Brazil | Canada | China | France | Hungary | Mexico | Spain | USA |
| No say | 46.1% | 42.0% | 50.2% | 60.1% | 35.0% | 12.2% | 16.7% | 42.7% |
| Some say | 27.8% | 22.2% | 17.0% | 23.3% | 21.6% | 29.4% | 27.8% | 21.7% |
| A lot of say | 12.9% | 8.8% | 18.9% | 11.9% | 8.4% | 26.7% | 11.2% | 11.4% |
| Complete say | 13.3% | 26.9% | 14.0% | 4.7% | 35.0% | 31.8% | 44.3% | 24.2% |

| James | Country | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Brazil | Canada | China | France | Hungary | Mexico | Spain | USA |
| No say | 45.5% | 46.5% | 42.7% | 60.0% | 30.9% | 11.7% | 17.3% | 42.3% |
| Some say | 28.7% | 21.7% | 16.4% | 21.6% | 29.3% | 27.5% | 25.2% | 33.1% |
| A lot of say | 15.0% | 13.0% | 28.0% | 13.6% | 13.2% | 26.4% | 14.0% | 9.6% |
| Complete say | 10.9% | 18.8% | 12.8% | 4.9% | 26.6% | 34.5% | 43.5% | 15.0% |

| Mary | Country | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Brazil | Canada | China | France | Hungary | Mexico | Spain | USA |
| No say | 35.8% | 13.5% | 7.9% | 21.4% | 14.5% | 7.9% | 13.7% | 16.7% |
| Some say | 19.0% | 10.4% | 10.6% | 12.6% | 10.2% | 21.9% | 22.2% | 15.9% |
| A lot of say | 18.4% | 13.1% | 26.8% | 29.7% | 10.0% | 36.0% | 12.9% | 14.4% |
| Complete say | 26.8% | 63.0% | 54.7% | 36.3% | 65.4% | 34.2% | 51.2% | 53.0% |

| Rita | Country | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Brazil | Canada | China | France | Hungary | Mexico | Spain | USA |
| No say | 34.7% | 27.6% | 20.7% | 39.3% | 33.5% | 8.8% | 14.0% | 31.9% |
| Some say | 37.4% | 41.1% | 23.1% | 41.6% | 27.3% | 26.0% | 35.8% | 39.6% |
| A lot of say | 14.4% | 13.4% | 38.9% | 15.4% | 12.9% | 29.4% | 12.5% | 11.3% |
| Complete say | 13.6% | 18.0% | 17.3% | 3.7% | 26.3% | 35.7% | 37.7% | 17.2% |

By just looking at these frequencies we may see that "Complete say" and "A lot of say" grades are rare with respondents from France. Coupled with the fact that the French people rated themselves very high, we may expect indeed that they will have greatest $\mu$ estimate. The situation is somewhat different with China. While their self assessment score was as high as that of France, we see that Chinese respondents tend to give considerably more "Complete say" and "A lot of say" ratings to all vignettes. This may mean that people in China have lower standards for what it means to them to control one's personal data. So we expect the model to correct for this and rank China's $\mu$ lower than France's. Another interesting change is expected with Mexico and Spain, which have a rather high self assessment score but give a lot of "Complete say" and "A lot of say" responses to persons in the vignettes.

## Fitting the model and discussion of results

The statistical model which we are going to use predicts one's amount of say $\mu$ and classification thresholds $\tau$ based on a group of variables which account for differences in cultural background. In our case we are mostly interested to carry out a cross-country comparison, but along with country we also include age, gender and race of the respondents. In a more elaborate analysis we may also wish to consider education, occupation and income as relevant background variables.

Here are the results of fitting a model with mentioned explanatory variables.

Figure 5: Regression coefficients for $\mu$

|  | "Amount of say" of respondents ($\mu$) | |
|---|---|---|
|  | Coefficient | Standard Error |
| gender | -0.0444 | 0.0363 |
| log-Age | -0.1298 | 0.0477 |
| white | 0.0377 | 0.0604 |
| Brazil | 0.2896 | 0.082 |
| Canada | -0.1768 | 0.0785 |
| France | 1.2311 | 0.0796 |
| Hungary | -0.5689 | 0.0798 |
| Mexico | -0.8071 | 0.0892 |
| Spain | -0.7737 | 0.0804 |
| China | 0.3485 | 0.084 |

As was mentioned above, the location and the units of the continuous scale on which we define $\mu$ have no meaning and are chosen arbitrarily. (The model is identified by having no intercept and fixing the variance of the random error to be 1). Thus the coefficients of the model can't be assigned their usual interpretation. However, we still can say which variables have significant contribution to the model and in which direction they influence the value of $\mu$. For example, gender and race (white indicator) don't significantly differ from zero; negative log-Age means that older people have, on average, less say; French people have, on average, more say than Americans (USA is the reference country because its indicator is not in the regression).

The four vignettes are assumed to have fixed values of $\theta$'s, whose estimates are

Figure 6: Amount of say in the vignettes ($\theta$'s)

|  | "Amount of say" of vignettes | |
|---|---|---|
|  | Value ($\theta$) | Standard Error |
| Mike (q29) | -1.0194 | 0.1986 |
| James (q30) | -0.9973 | 0.1979 |
| Mary (q31) | 0.4048 | 0.1993 |
| Rita (q32) | -0.6858 | 0.1972 |

It is difficult to see from this table how the vignettes compare to the respondents, but we clearly see how the vignettes are ordered. In particular we may notice that Mike

and James are almost identical, so the analysis results would not change if one of these two vignettes was missing. However, the situation may change if we add more background variables to the model.

The last component of the model is the thresholds, which people use to categorize their perception of **μ**.

Figure 7: Regression coefficients for thresholds (τ)

|  | Thresholds | | |
|---|---|---|---|
|  | No / Some | Some / A lot | A lot / Complete |
| Intercept | -3.113 | -0.5875 | 0.9115 |
| gender | -0.0783 | -0.0371 | -0.0281 |
| log-Age | 0.5052 | 0.165 | -0.0858 |
| white | -0.1164 | -0.0607 | -0.0174 |
| Brazil | 0.5096 | 0.204 | 0.2775 |
| Canada | -0.0608 | -0.2231 | -0.1538 |
| France | 0.6496 | 0.4171 | 0.6749 |
| Hungary | -0.0928 | -0.4522 | -0.6013 |
| Mexico | -1.1795 | -1.1087 | -0.8138 |
| Spain | -0.8199 | -0.8097 | -0.9778 |
| China | -0.0665 | -0.7084 | 0.0171 |

This table can be useful for comparing standards of classification between different strata. For example, white people have a bit lower standards than non-white; Mexicans and Spanish have considerably lower standards than Americans (that is, if an average American judges someone to have "A lot of say", an average Mexican may classify as "Complete say").

To visualize the estimated mean "amount of say" between countries, we fix other parameters of the model, thus picking a reference group, and calculate **μ**'s from the model for that reference group across countries. We can also calculate the thresholds for that reference group and see how each country would be classified by an average representative of the reference group. Note that it is not important how we choose this group when we compare just the respondents, since it will only affect the location of estimated **μ**'s but not the differences between them. If we want to compare respondents with vignettes, the reference group is important because vignettes are fixed values.

The chart below uses the following reference group: Age = 30 year old, race= white, gender = male. The thresholds shown are those of the USA part of the reference group.

**How much say do you have in what happens to your private information?**
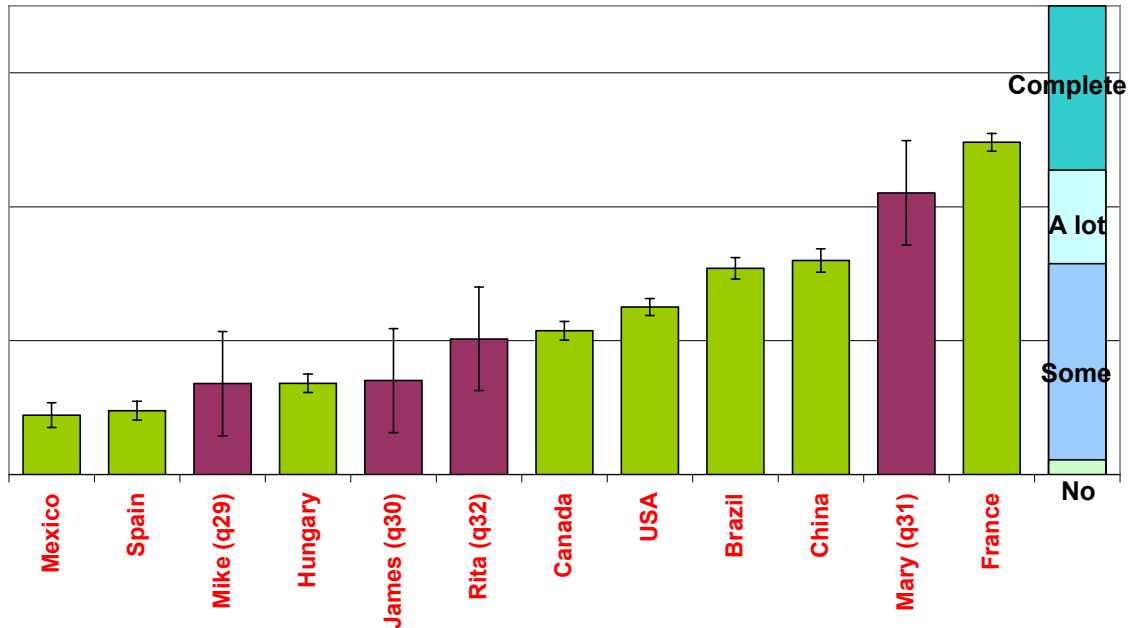


We see that, as was intuitively expected in the preliminary analysis, France remains at the head of the list. It is still followed by China, but its difference with France has changed from none to quite big. Another big change has occurred with regard to Mexico, which exhibited very low standards for having much say, and we now see it at the tail of our ranking. The same situation applies to Spain. We still observe that all countries would be classified as having at least "Some say", however we now have to add that this is according to American standards (thresholds). If we picked extreme thresholds, like those of Mexico or France, we might see a very different classification picture.

**Which responses to use?**

After fitting the vignettes model we can calculate the estimated "amount of say" **μ** for each respondent. These values are supposed to provide a score on how much say respondents have in what happens to their private information, consistent in the cross-cultural context. The reader may now wonder if he or she should use the original or corrected responses in any subsequent analysis involving Q2. It also becomes a point of argument whether the other questions, where vignettes were not available, provide a valid picture of the concepts they address?

The answer is that both original and corrected responses may be appropriate to use, depending on what interpretation the results are given. When we use raw responses, we really only work with people's opinion, and therefore all findings should be stated

accordingly. For example, we have seen that 18.4% of 1071 Mexicans reported having "Complete say" in what happens to their personal data. It would be incorrect to infer from this that 18.4±0.47% in the population of Mexico have "Complete say"; this is rather an estimate of how many people in Mexico would rank themselves into this category. After we have reanalyzed the responses with the vignettes, we can obtain objective results. For example, we can quantify the difference in the "amount of say" between Mexico and other countries or compute correlation of "amount of say" with other variables.


## Appendix: Vignette questions

**Q29**: [Mike] goes to the drug store to buy film, which was advertised to be on sale. He finds out at the store that in order to receive the discount, he must apply for a customer loyalty card, which involves filling out an application form. It requires [Mike] to fill out his home address, occupation, and marital status. He fills the form out to get the special pricing. To what extent does [Mike] have a say in what happens to his personal information?

**Q30**: As part of the concern for national security, assume that the government creates a database to search for terrorist activity. All government records are merged with any available commercial data such as bank records, credit statements, and travel manifests. Citizens, such as [James] are required to provide fingerprints, photographs, and iris scans. [James] does this. To what extent does James have a say in what happens to his personal information?

**Q31**: [Mary] pays cash at a large, crowded department store and provides no information about herself to the cashier. The cashier asks for [Mary's] postal code/zip code; [Mary] refuses and still makes her purchase. To what extent does [Mary] have a say in what happens to her personal information?

**Q32**: Assume that as part of the concern for national security, the government creates a database to search for terrorist activity. The government requires everyone, including [Rita], to submit an annual form containing detailed information about themselves. The form asks for employment information, criminal activity, and any travel abroad by the individual or any family members in the last five years. [Rita] complies, but decides not complete the section on travel. To what extent does [Rita] have a say in what happens to her personal information?

References
King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon, "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research," American Political Science Review, 98(1), 191-207, 2004.