

Representative, or did you mean representative, or maybe representative?

By

Created 2007-10-22 00:50

Representative, or did you mean representative, or maybe *representative*?

The question of representation has been discussed a great deal in openDemocracy's coverage of Tomorrow's Europe poll. There is confusion of terminology - statistical representativeness does not mean political representativeness, does not mean experimental representativeness. There is also a surprising amount of theoretical contention in both the statistical and political senses to make the opportunity for confusing conflation truly vast.

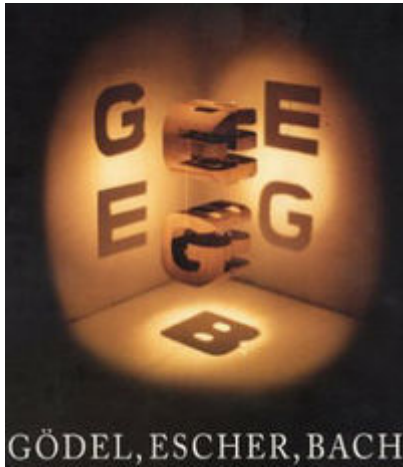
Clive has asked whether the stratified sampling used on Tomorrow's Europe can be representative when the selection was biased in order to include more of certain nationalities than would have normally come out of a random sample of this relative size. The DP designers handled this problem in the way discussed by Fishkin ([here](#) [0]).

Taking it as given that practicalities restricted the number of final participants to approximately 400, a random sample would be unlikely to contain many representatives of the smaller nations. Yet one of the hypotheses being tested is that when Europeans come together to deliberate, their views are changed. How can we reconcile random selection of participants with the desire to have "Europe speak"? This, of course, is as real a problem in the representative institutions as in this Deliberative Poll.

The DP organisers chose a course which seems very sensible: they constrained the sampling to reproduce approximately the European Parliament's distribution of voices. In some sense, the conversation being modelled was that found in the environment of the Parliament.

Very sensible, although it poses a few questions from the statistical view of representation. The first, which Fishkin writes about, is that inferences cannot be made about the influence of nationality on outcomes without a prohibitively large sample. So if one believed that nationality was the main, or even a major, determinant of views or of one's reactions to new information, this would indeed be trouble. (Actually, the results could presumably be aggregated somewhat - "Southern Countries", "Countries with strong Euro-sceptic parties", "Lutheran majorities" etc - to test some geographical hypotheses with sufficient number of observations.)

The second problem arises if there is correlation of any kind between Nationality and any of the other factors being controlled for (age, occupation, education levels etc). The Maltese, for example, are over-represented relative to their population share. What if they also tend to be younger than most? The correction for nationality will bias the other variables we would like to keep neutral - privileging the Maltese, in our example, also necessarily privileges the young. The nature of the samplers' problem can nicely be visualised with the "GEB" projector on the cover of the book "Godel, Escher and Bach".



Look at the wood carving in the middle of the picture that must accurately project three different letters when light is shone-through from three different faces. It is a complex and intricate object. Now notice that the carving you make in one plane affects the carving in another plane - that the two are correlated. There is now a real problem in trying to get perfect, independent projections from the three sides--it is not obvious that this is even always possible. One solution for correlation effects like this is to take very large samples.

A statistically attractive alternative to the stratified random sampling would be to get a large enough sample of individuals for each variable whose significance we wanted to test. This is the scientist's option: make sure we have 40 Maltese; make sure we have 40, or 100 people aged 70 plus, etc. If we wanted to then produce a "representation" of what happens Europe-wide under deliberation, then we would just scale our results by the appropriate proportions.

But the statistical sense of representation conflicts here with both the experimental and the political sense. Fishkin is looking at the impact of an Athenian-style discussion in a microrcosmic Europe. The discussion with 40 Maltese and 40 oldies etc. would be nothing like the discursive dynamic where the proportions reflect actual population proportions. The notion of political representation demands that population proportions be preserved. The experimental simulation requires it also. But the statistical would require an uneconomically large sample to also satisfy that aspect of representation.

Tomorrow's Europe clearly had a tough balancing act between statistics, politics and budgets. The decisions made seem extremely sensible and pragmatic. I look forward to having all the data being made public so that unexpected correlations can be checked, so that the significance of the groupings controlled for can be examined ex post, and so that different aggregations - like looking at South/North, Catholic/Protestant etc - can be mined for insights.

(With thanks to Graeme Mitchison for sorting me out on some statistics (amongst other things))

Source URL:

http://www.opendemocracy.net/blog/dliberation/representative_or_did_you_mean_representative_or_maybe_representative