A Brief Response by A. Shaked to Fehr & Schmidt's Reply to Shaked's Pamphlet

8 March 2005

Fehr & Schmidt's reply to my pamphlet is short and succinct. They point out two imprecisions in my pamphlet[1]. Fortunately, these do not alter any of the main arguments made in it.
In their reply, Fehr and Schmidt do not address the central issue of the *Rhetoric and Methodology of their Calibration Exercise*, which the pamphlet raises.

How well founded is the calibration method practiced in their paper?

Imagine you have a theory and available data for a situation *A*, possibly an experiment, for which the theory applies.

You declare your wish to calibrate the theory with the data of situation *A*, and explain with it the data of another situation *B*. Situation *B* may be thought of as a single or a number of situations for which the data is already available or will be available in the future.

You apply the theory to situation *A*, and you find a host of parameter distributions that are compatible with the data. How do you pick up a single distribution from among this large set of distributions? You have no pre-set rules that prescribe how you should choose a single distribution. However objective and well meaning you may be, you may be influenced by the data of *B*, which you may have already seen at this stage.

You find that in order to explain the data of some situation *B* a certain parameter (call it β) has to be further restricted to a certain range. You adjust the chosen distribution, or if the data for *B* is already available at the beginning of this exercise, you take this restriction into account when you choose the distribution. Alternatively, you declare the β's outside the 'right' range to be *too extreme* for the theory. Your original theory allows the parameter β to take these values. You make this condition on β only *after* you discovered that explaining situation *B* requires it.

For another situation *B* you find that in order to explain the data, the parameter β should be correlated with another parameter (call it α). You adjust the distribution of your parameters to allow for the correlation.

All too often, the absence of methodological foundation is accompanied by rhetorical devices. The effect of these devices is that they help conceal the true nature of your methods and at the same time they enhance the message of your paper. You do not

[1] These have now been corrected; the new version of the pamphlet can be downloaded from http://www.wiwi.uni-bonn.de/shaked/rhetoric/

openly tell your readers how you chose your distribution, and how this distribution was affected by the data of *B*. You place significant details in footnotes and appendices, where few readers will ever notice them. When you make your computations, you do not stress the influence the data of *B* had on your choice of parameters. After some time elapses without anyone openly criticizing your methods and computation, you refer to this computation, in your future papers, as an *accurate* explanation of *B*. Now you are set, the profession is likely to remember only your latest references to your calculations and no one questions your statements.

Can you honestly say that you have explained the behavior in situation *B* with a calibration using the data of situation *A*? You have used the data of the *two* situations *A* and *B* to help you choose a particular distribution of parameters. In what way was the data of *B* less instrumental in the choice of the distribution than the data of *A* ? Unless you clearly define degrees of intensity of use of data, it cannot be said that the data of *A* was instrumental in the calibration, while that of *B* was not. You have therefore failed to explain the data of *B*.
What you may have done is to calibrate your theory using *both* data sets of *A* and *B*. You may have succeeded in finding one distribution of parameters that agrees with the data of *A* and *B*, but you have not explained any data at all. By tying your hands and keeping this distribution constant you may now begin to explain new situations and data as they come along.

Irrespective of how many researchers, papers or even books apply such methods; can these tools be called rigorous and scientific? What scientific theory supports and justifies such methods?

The least that this method requires in order to be called scientific is a complete separation of its two elements: That of choosing the distribution with the help of the data of *A*, and that of explaining the data of *B*. There should be clear rules, set in advance, defining how to pick up a single compatible distribution from the given data (of *A*). This process should be completely independent of any situation B and its data.

It is these methodological problems that should be discussed before practicing these calibration techniques.