SCIENCE SPEAKS ONLY ONE LANGUAGE

by

Ray Pawson
Department of Sociology
University of Leeds
Leeds, U.K.

Discussion Paper

on

Claes Fornell's

THE BLENDING OF THEORETICAL AND EMPIRICAL KNOWLEDGE IN STRUCTURAL EQUATIONS WITH UNOBSERVABLES

The Thirteenth International Conference on the Unity of the Sciences
Washington, D.C. September 2-5, 1984

SCIENCE SPEAKS ONLY ONE LANGUAGE

Let me begin with two central points of agreement with Fornell's paper. In his attempt to find methods which blend theoretical and empirical knowledge he puts his finger on what is undoubtedly the central dilemma in all social science research, namely our need to come to a better understanding of the consequences of the interdependence of theory and observation. Furthermore, he is absolutely correct in charging that much of what passes for sound practice in social science data analysis is based on epistemological assumptions which obtained in the dark ages of the philosophy of science. In these respects what Fornell has to say needs to be shouted from the academic roof-tops - refinement in measurement and data analysis is redundant, indeed impossible in isolation from theoretical development.

The Double-Language Model

Where I disagree with Fornell is in terms of the manner of the impregnation of theory in data and some of the ramifications of their interdependence. In a nutshell my objection is that Fornell still utilizes what is referred to in the philosophical literature as a double-language model for theory and observation, a formulation which is most evident in his separate treatment of 'abstract' and 'empirical' meaning and the discussion of the 'directionality' of the relationship between theory and data. Now the double-language model represents a huge advance over empiricism in its recognition that careful observation of facts

'out there' is not the be all and end all of science and quite properly insists we acknowledge the role of theoretical discourse with its attendant use of abstractions, metaphors, models, mechanisms and so forth. However, in the last decade or two, philosophical orthodoxy (if there be such a thing) has turned to the view that no useful distinction can be maintained between observational and theoretical terms and concluded that science does in fact speak with only one language.

The two-language model has been much criticised on the grounds that no satisfactory account of the status of correspondence rules has been formulated, the problem being that as soon as one allows for the interdependence of theory and observation, it seems pointless to refer to them as independent realms and as a consequence all accounts of the translation from theoretical to observational statements end up by proclaiming the priority of one or the other. For instance suppose one begins by trying to identify how the realm of observation is constituted. The obvious difficulty is that sensory experiences depend on the observer having certain general hypotheses about the nature of the thing observed. Recall Popper's famous invitation to his positivistically inclined students to make observations of their lecture room - the diversity of observational reports manifests their dependence on prior expectations of what is relevant. Recall also the plethora of illustrations from the psychology of perception (duck/rabbits, the dot and frame illusion etc.) which show how even apparently direct sensory perceptions depend on or indeed <u>are</u> hypothetical judgements.

Starting from the opposite end and supposing we were attempting to decide between competing theoretical propositions on an issue. The double language model would have us adjudicate on the basis of the observational consequences that flow from the theoretical propositions via correspondence rules. But allowing for the interdependence of theory and observation obviously means that the selection of correspondence rules is itself made according to one's preferred theoretical postulates. Thus evidential assessment of theoretical propositions turns out to be no assessment at all. I cannot rehearse any further objections to the two-language model here [Papineau, 1979, Ch. 1], suffice it to say that the only way out of the circularity displayed in the logic above has been to resort to a position which asserts the ultimate priority of one or other of the two realms. Note in this respect, Fornell's (p. 15) remark that in structural equation models, '.... the theoretical variables are required to be composed of nothing but a combination of empirical variables. This assures that the analysis cannot go "beyond the data" .....'. This brings us to the central dilemma concerning statistical modelling in social science, that is to what extent do they allow a genuine role for theory in their portrayal of the world?

Post-Empiricist Philosophy

However we are not yet in a position to answer the question until we know of some alternatives to the double-language model. Post-empiricist philosophy has in recent times produced some revised interpretations of

the relationship between theory and evidence which can be briefly described under two or three headings as follows. First to go is the distinction between two languages for concepts and observables [Achinstein, 1968, ch.5, Feyerabend, 1975, ch.6,7.], instead all terms in science are said to take their meaning discursively or wholistically according to their place in relationships, definitions and laws, much as in the manner that Fornell describes in his section on 'abstract meaning'. However according to the newer philosophies, evidence, or what Fornell calls 'empirical meaning', is constructed in exactly the same wholistic fashion, that is to say evidence is constructed by understanding how a property is locked into a system of relationships rather than direct perception of the property or of its so-called indicators.

All this, of course, would come as no surprise in the natural sciences, where this more creative aspect of measurement is commonplace. To take the elementary example of the galvanometer: clearly in discussing its operation we are not dealing with direct and untainted observation since the very rationale behind the instrument is a theory of, and an ability to marshal, electromagnetic forces. We can report that the galvanometer needle moves but to interpret that movement needs an understanding of motion in electromagnetic fields. What is more, the actual construction of the instrument calls on a host of secondary theories - the galvanometer needle is set on jewelled bearing so as to minimize friction, it is as light as possible to deal with the effects of inertia, the

instrument is encased to minimize the influence of extraneous magnetic forces and so forth. In short the evidence drawn from such sources is empirical in the sense that it involves some manipulation, some intervention in the world but not in the sense that the evidence is directly experienced and thus can be considered to have epistemological priority.

Another recently emphasised feature of the meaning of scientific terms is their derivation from generative mechanisms which produce the relatively enduring regularities which structure the world. [Harré and Madden, 1975, Bhaskar, 1975, c.f. Fornell, p.8]. Hence, unlike much statistical modelling in social science, explanatory requirements are not satisfied simply by unearthing patterns and regularities in the occurrence of events, but by devising an understanding of the underlying mechanism that generates and constitutes the causal regularity. Again the idea is that scientific concepts should not be thought of as separate, singular objects or variables each answerable to its own element of reality. Rather we know the meaning of a concept (or synonomously, the parameters of a measure) by constructing models and analogies of the internal structures of physical (or social) systems, so as to show how a system passes from one state to another and it is this understanding which informs us of the relationship between the component properties of the system and thus provides knowledge of the numerical characteristics of those properties.

The time-honoured example of this generative derivation of measurement properties is the establishment of the absolute temperature scale. The kinetic theory of gases depends upon the analogy of a swarm of microscopic particles moving in a confined space and uses the principles of classical mechanics to describe their motion and in so doing derives expressions for the temperature (average kinetic energy of the molecules) and other properties of a gas. The mode of reasoning tells us what temperature is, allows us to construe it as having a zero value and thus assign a metric scale, and finally to predict how it varies with other properties. Again the meaning is not derived from observation (since zero temperature has, of course, no empirical equivalent) and in fact the kinetic model supplies the parameters of temperature in advance of observation, the experimentalist being left to invent those instruments which best exemplify the idea.

One further feature of the post-empiricist accounts worthy of our attention here, draws on the idea that the basic units of scientific analysis are not concepts and observables but whole networks of theory and evidence [Hesse, 1974]. That is to say explanation deals with wholes or structures and it is these research programmes in their entirety which are evaluated in the course of scientific debate [Lakatos, 1970]. Pursuing the network metaphor further enables us to see the growth of science as the business of ever enlarging the numbers of threads converging on our existing stock of knowledge. The idea is that either experimentally or by deduction new concepts and new observations are grafted back into the

network. The result is that some 'knot concepts' or 'systematic measures' come to have a key coordinating role in science, they are easily identified by many theories and many instruments. [Kyburg, 1984].

Again one must draw a contrast between such programmatic thinking and statistical modelling in social science which operates much more at the level of specific regularities and tends to draft in theory to explain such correlations on a rather ad hoc basis. The central question perhaps boils down to whether the notions of 'unobservables', 'multiple indicators' and so forth represents an effort to move towards, or simply an effort to disguise the lack of, these network of coordinated reasoning which are so vital to measurement in science.

Explanation and Structural Equation Modelling

I have argued for a revised understanding of the interdependence of theory and measurement which can be no better summarised than in Kuhn's words as follows:

> "The laws of nature are so very seldom discovered simply by inspecting the results of measurements made without advance knowledge of these laws. Because most scientific laws have so few quantitative points of contact with nature, because investigations of those contact points usually demand such laborious instrumentation and approximation, and because nature itself needs to be forced to yield the appropriate results, the route from theory or law to measurement can almost never be travelled backwards. Numbers gathered without some knowledge of the regularity to be expected almost never speak for themselves. Almost certainly they remain just numbers'
>
> [Kuhn, 1961, p. 174]

This is the formulation which provides the authentic test of the aspirations of structural equation modelling with multiple indicators. To put the matter most starkly, measurement practice should be seen as an artefact of the formal networks of coordinated reason which constitute normal science explanation. Does such a situation apply in social science?

It is safe to say that social science in general in its empirical work has not operated in the manner described above. Rather we derive measurement parameters in ordinary language usage; that is we utilize the counts, orderings and categories available in everyday descriptions of the topics in question. [Cicourel, 1964]. Examples of this will be familiar in every substantive area: sociologists use 'number of years of schooling' to indicate educational attainment, economists use 'market share held by x firms' to indicate monopoly power, psychologists use 'ratio of test items answered correctly 'to indicate intelligence and so forth. To adapt the idea to Fornell's terminology the problem is that we attempt empirical research in social science when abstract specification of meaning is terribly premature. We measure in the absence of the potential richness of meaning that he documents; definitional specification often goes no further than claiming the indicator is an 'aspect of' some broader theoretical construct and specification by antecedents and consequences seldom tells us more than to expect that there will be a (positive or negative) relationship between X and Y. In other words it is because our explanations are not wholistic, generative, programmatic, network-like

and so on that allows or rather demands the creation of arbitrary, disconnected, first-order indicators to stand as our measures.

It is precisely this state of affairs which leads to the retention of the double language model in structural equation modelling. To utilize one of Fornell's own examples, one notes that in the absence of detailed knowledge of the mechanism whereby monopolistic power is achieved that we have no choice to take our chance with an array of competing indicators. Similarly, if our understanding of the consequences of monopoly power allow us to venture no further than to 'predict a negative relationship between monopoly power and exit and a positive relationship between monopoly power and voice', then, within certain limits of plausibility, anything goes in terms of selecting indicators to confirm such findings. In short, it is theoretical indecision which allows the observational level to live on in the form of common sense descriptions of the objects of investigation.

The first consequence of sticking with the double language model is the ultimate preference, mentioned above, for one or other of the theoretical and empirical domains in the production of results. I have argued elsewhere that, in general in structural equation models with multiple indicators, the so-called theoretical constructs merely summarize observations made and the only important difference between the various estimation procedures is the method chosen to combine and summarize the information. [Pawson, 1980]. Thus arithmetically speaking no such models can go 'beyond the data', the basic content of empirical generalizations is,

after all, empirical information.

The important question that Fornell raises is the possibility that different estimation techniques (covariance analysis and partial least squares) can change the balance of observational and theoretical input. We need to be enormously careful in interpreting such a claim. What changes between the techniques as described by Fornell is not a reformulation of measurement practice, we fire-off the same measurements and theory enters at a later stage to adjudge the veracity of the findings. So instead of theory informing us how to measure as in the network model, we measure anyway and the job of theory is to gauge how impressed we are by the results.

Hence the preference for formative or reflective indicators is not a choice which allows theory into the fabric of measurement decisions, it is a device which operates at the level of 'explanatory power' of a theory. And typically for the double language model, the concept of explanatory power of a theory turns out to favour a purely empirical connotation in terms of 'variance explained'. Thus the scope for theorizing in these models boils down to a choice between a decision to affirm an empirical generalization on the basis of what one acknowledges are low quality indicators or to maintain more qualified support for a generalization on the basis of what one assumes are high quality measures. Thus, one is asking theory to act an extremely blunt instrument, which involves less the matter of explaining how a relationship comes about and more the matter of expressing faith in its significance. I fail to see that there can be any

formulation of the general conditions for operating the proposed trade-off between indicator validity and explanatory power; the decision between them must rest, as Fornell admits, with knowledge of the contexts which pertain to each regularity studied. However, to repeat the whole point of my note, if one has detailed knowledge of the processes underlying a relationship then this involves knowledge of the constituative properties in the first place. Knowledge of the mechanism brings knowledge of the measure. If, by contrast, ones understanding of a particular context is so flimsy that one is left with declarations of faith in a generalization or faith in a measure, then one has no business to be performing exact estimates of the said relationship in the first place.

Conclusion

Let me try to summarize my points of agreement and disagreement with Fornell. Starting from exactly the same premise that percepts without concepts are blind and proceeding with exactly the same instinct that the processes of concept formation, theory construction, measurement and hypothesis testing are inseparable, we appear to disagree only on the potentialities of the various statistical estimation techniques. In so far as Fornell is saying that the discrepencies in the estimates under P.L.S and covariance models, reflective and formative indicators and so forth demonstrate the plasticity of the relationship between theory and data, I can only shout hear-hear! If, however, he is claiming a solution and trying to champion one particular statistical estimation technique as

somehow providing the authentic balance between the conceptual and the empirical, we totally disagree. Science, I would venture, progresses by delving into the mechanisms that constitute laws and regularities and not by opting out of the inner workings of relationships by settling for the safety of statistical estimates of them.

REFERENCES

Achinstein, P. (1968) — Concepts of Science, Baltimore : John Hopkins Press

Bhaskar, R. (1975) — A Realist Theory of Science, Leeds : Leeds Books

Cicourel, A. (1964) — Method and Measurement in Sociology, New York : Free Press

Feyerabend, P. (1975) — Against Method, London : New Left Books

Harré, P and Madden, E. (1975) — Causal Powers, Oxford : Blackwell

Hesse, M. (1974) — The Structure of Scientific Inference, London : Macmillan

Kuhn, T. (1961) — 'The Function of Measurement in Modern Physical Science' in H. Woolf (ed.), Quantification, Indianapolis : Bobbs-Merrill

Kyburg, H. (1984) — Theory and Measurement, Cambridge : C.U.P.

Lakatos, I. (1970) — 'Falsificationism and the Methodology of Scientific Research Programmes' in I. Lakas and A. Musgrave (eds.), Criticism and the Growth of Knowledge, Cambridge : C.U.P.

Papineau, D. (1979) — Theory and Meaning, Oxford : Clarendon Press

Pawson, R. (1980) — 'Empiricist Measurement Strategies', Quality and Quantity, 14, pp 651-678.