Jan-Erik Lane
Department of Political Science,
University of Umea, Sweden

The Theory/Concept/Indicator Problem: An Epistemological Note for COCTA

Lane, J.-E.: **The theory/concept/indicator problem: An epistemological note for COCTA.**
In: Int. Classif. 15 (1988) No. 1, p. 21–24, 18 refs.

Social science knowledge has two basic modes: theory or concepts on the one hand and data or indicators on the other hand. How do we strike a balance between the two? The positivist philosophy of science overemphasized the empirical element whereas the new theory of science outlined by the adherents of Kuhn and Feyerabend state too strong a case for theory (theoreticism). It is equally wrong to reduce concepts to data as it is to reduce data to theory. Concepts and indicators should be stated in two different languages so that we may strike a balance between the two. Theory without data is empty and data without theory is blind. (Author)

1. Introduction

Since the Second World War the social sciences have become more conscious about their epistemic preconditions. There has been a search for adequate concepts and substantial evidence, i.e. the establishment of data sets as wide and as representative as possible in combination with an ambition to do away with particularistic and culture-bound conceptions. We find an elaborated debate about the principal methods for obtaining empirical information for social analysis – the case study, the comparative method and the statistical approach (Przeworski & Teune, 1970; Lijphart, 1971; Lazarsfeld, Pasanella & Rosenberg, 1972). Moreover, the importance and relevance of theory for the interpretation of the data has been established as a result of the desire to make the social sciences truly systematic.

It has been taken for granted that first comes the construction of a theory or an approach; then the concepts are operationalized in terms of indicators which lend themselves to the statement of the data in terms of which the theory is eventually tested – confirmed or rejected. The outcome of this lack of reflection on the interaction between theory and data has been a proliferation of theoretical frameworks and a variety of indicators which defy the idea of a cumulative growth in knowledge (Sjöblom, 1976). It is almost as if each scholar started from scratch, constructing an idiosyncratic framework in relation to idiosyncratic indicators. The social sciences run the risk of developing as many towers of Babel as there are innovative and independent social science scholars (Sartori, Riggs & Reune, 1975). Instead of cumulative growth in comparative knowledge there will be a state of theoretical solipsism, each scholar creating his/her own world. Several social science methodologists argue today that such a predicament is not only the actual state of affairs in the social sciences, but that it is the only available option. I will discuss the new social science methodology – theoreticism – in order to state a case for a balanced mix between theory and data.

2. The New Methodology

According to the new methodology we are to consider that theory is more basic than data and concepts are more important than evidence. The new methodology expounded by Feyerabend (1975) and Kuhn (1962) among others rejects the empiricism of logical positivism and the falsificationism of the Popper school; instead it argues in favour of a kind of theoreticism that implies that the theoretical structure of science governs its empirical elements to such an extent that an empirical test of theoretical propositions is impossible and that the idea of an independent set of empirical data is indefensible. Although theoreticism is at odds with the common sense notion that statements about entities like legitimacy, alienation or political stability may fail to satisfy a body of evidence or data, it is no exaggeration to maintain that theoreticism is the methodology of today, at least within the social sciences (Morgan, 1983).

Yet, the methodological idea of empirical confirmation of theoretical statements is a fundamental presupposition in social inquiry. Indeed, it is difficult to do away with the belief that a criterion of adequacy for social science knowledge is the extent to which theory satisfies evidence or abstract statements are confirmed by a set of data. It may be the case that the distinctions between theory and data as well as between concepts and indicators are anything but clear, but it still seems that theoreticism is an extreme solution to the problem of evaluating social science theories. It appears to be difficult not to say impossible to find a solution to basic social science puzzles without some kind of empirical test that guides us to the choice between alternative problem solutions.

3. Data and Reality

To the social science methodologist the distinction between proposition or hypothesis and reality does not appear troublesome. Analyzing various types of social systems the social scientist is bound to accept the common sense notion that the assertion itself may be separated from the subject of the assertion. Statements have properties like true or false, confirmed or disproved, precise or imprecise, that social systems do not have. Whereas the distinction between knowledge and reality is not fundamentally ambiguous, the separation between theory and data is much more troublesome. Theoreticism maintains not only that theory determines concepts (Sartori, 1984), but also that data is a function of theory. However, if data is used to evaluate alternative theories, then how can we choose between theories? Theoreticism introduces a vicious circle – theory determines the data used to evaluate the theory.

In relation to the methodology of the social sciences it may prove clarifying to separate two distinctions, the one between knowledge and reality and the other between theory and data. The problem of identifying how

theory and data interact is different from the question whether there exists an external world outside of human knowledge. Thus, we have a 2 × 2 Table:

Table 1. Theoreticism versus Empiricism

EPISTEMOLOGY	ONTOLOGY: Knowledge	ONTOLOGY: Reality
Theory	I	II
Data	III	IV

Distinction B (Theory / Data)

Distinction A (Knowledge / Reality)

That beliefs in social science inquiry are true in the sense of correspondence to reality (distinction A) is not so difficult to accept in a social methodology. But the methodology (distinction B) that states that a set of empirical terms may be separated from a set of theoretical terms allowing the introduction of the data and theory or indicator and concept distinctions is troublesome. How could there be a body of terms the use of which requires no theory at all?

4. Positivist Metaphysics

Evidently, the social scientist believes that theories are true of reality because they agree with social systems properties (II); and the extent to which beliefs agree with reality or correspond to reality can only be judged by the extent to which beliefs are in agreement with data about reality (III). Since such data has to be reported in propositions of some sort the requirement of a correspondence between theory and reality becomes an insistence upon a relationship between some kinds of propositions labelled "theory" and other types of propositions labelled "data". Since we cannot grasp reality directly, the evaluation of theory cannot be made on the basis of a comparison with reality. And consequently we enter the vicious circle of theoreticism: if data basically are nothing but statements about reality, then how can we assume that data are more true or basic than the theories that are to be evaluated with reference to the data? If data are not reality but express beliefs or assumptions about reality (IV), then how can we distinguish between data and theory? Instead of comparing theory and reality in order to establish the degree of correspondence we are in fact comparing theories, since data is theory-loaded (I). However, once the positivist methodology of a double language is abandoned the conclusions of theoreticism do not follow. From the fact that there is no absolute distinction between theory and data or concept and indicator allowing us to identify a basic empirical language that is not open to doubt, it does not follow that a relative distinction may not be made between beliefs that are problematic and beliefs which are unproblematic. Both beliefs may be theoretical in the sense that they contain or assume certain theories. In order to establish the truth claims for social science theories that compete for our allegiance we use theories which we are in agreement about. These unanimity beliefs establish the data and define the measurement procedures expressed in the employment of indicators. In order to make general statements we employ uncontested theories in order to arrive at a body of data corresponding as closely as possible to reality as well as a set of indicators which allow the measurement of the abstract concepts in the contested theories in relation to a body of data.

5. Theory and Data Reciprocity

The relationship between theory and data and concept and indicator is a many-to-many relation meaning that theory is *empty* without data: if a new theory has no implications concerning the beliefs we adhere to, then why should we bother? Actually, if there were no relationship between theoretical concepts and accepted notions about data, then what would be the content of such theories? If theory did not explain data, then would not theory just be a word game? Since the relation is not a one-to-one relation the same indicators and the same data are relevant to alternative theories, which means that empirical evaluation is possible. On the other hand, data without theory is *blind*, because beliefs about data and indicators do not allow the mechanical derivation of theory. The implications of indicators and data have to be sought out as one and the same body of data allows the construction of different theories.

Theoreticism is the overemphasis of one of the elements in the interaction and reciprocity between theory and data, concept and indicator. The dynamic and open-ended aspects of this relationship mean that each element needs the other for its development. It is just as important to emphasize that the data may admit alternative interpretations as that theories may satisfy certain data but not others. Is it possible to substantiate such a *reciprocity methodology* about the interaction between theory and data in the social sciences? Whatever way one integrates the two sides of knowledge – theory and data – in the social sciences it is impossible to bypass the fundamental place of the indicator and its employment in order to arrive at data bases. What, then, is an indicator?

6. Indicators as Empiricism Reconsidered

Presumably, the attempt to clarify the nature of an indicator or an index is based on the assumption that a distinction may be made between a construct and an indicator. Following this assumption we may ask for the relationship between the two, by means of which an indicator is said to indicate or measure a construct. A priori, there seems to be three possibilities: convention, analyticity and causality. F.S.C. Northrop argues in favour of the solution that indicators are tied to constructs by means of convention (1947). Northrop distinguished between "concepts by postulation" (theoretical terms) and "concepts by intuition" (indicators) and states that whether an indicator indicates a construct is a matter of agreement. Blalock also argues in favour of the consensus solution (1972). The difficulty with the convention solution is that it is not quite clear what exactly the statement that consensus is the basic tie between construct and indicator refers to. *When* was this convention estab-

lished? By *Whom?* Recorded *Where?* It seems somewhat peculiar to justify the use of an indicator to measure a construct in social inquiry simply by stating that such is the convention. Obviously, the rationale must be something more basic than consensus as we all know that consensus is hard to come by and easy to break. Moreover, stating that there is a convention behind the linkage is simply a petitio principii, because the convention must have a rationale other than simply the fact that it was so agreed.

A causal interpretation has been suggested (Blalock, 1971; Blalock, 1974). The difference between a causal interpretation and a consensus interpretation may be noted. Conventions are made by man to be broken if not upheld; causal relationships are there to be discovered and once discovered they cannot be undone. The causal approach to the problem of the nature of the indicator is common and it is usually stated in terms of a relationship between latent and manifest variables. However, even in measurement models where the relationship between indicator and construct is estimated by means of path analysis one may find a distinction between the causal relationship of latent variables and so-called "empistemic correlations" between the indicator(s) and the construct (Sullivan & Feldman, 1979). The expression "epistemic correlation" seems a curious blend of two alternative ways of interpreting indicators, as knowledge conventions or as statements that may be true or false.

A third interpretation, the analytical interpretation, is conceivable. The relationship between indicator(s) and constructs may be conceived as a relationship not between two different entities but as a relationship concerning one and the same entity. An indicator is "an aspect of" or a "part of" an underlying variable (Blalock, 1971) or a set of indicators may be combined into an index that serves as a summary measure for the attribute space of a construct (Lazarsfeld & Rosenberg, 1955). Indicators tap concepts, indicators identify various aspects of a concept; consequently, they are not causes or effects of concepts, because concepts do not have causes or effects, only objects have, and a whole can hardly be conceived of as causing its various "aspects" or "parts". Nor do summary variables cause anything; they are measures of objects that have causes and effects.

Although the convention and causality interpretations may have their advantages it is difficult to see that the analytical interpretation clarifies much. If it is a matter of deriving the indicators from a construct by means of logical analysis, then why is it that the identification problem is acute? If indicators may be derived by conceptual analysis, then of what use are the various techniques employed to estimate the construct by means of a set of indicators? Typically, we do not know one of the elements in the construct-indicator interaction, but the analytic solution claims that the one element entails the other. It seems safe to exclude the idea that the relationship between indicator and construct is a logical relationship. We are left with two options: convention or causality. And the restriction on the choice of either one of these options for the construction of a methodology is that we must allow for the possibility that theory and data may interact so as to strike a balance between the two basic epistemic elements.

7. The Balance Methodology

A major trend in social science methodology favours the idea that theory is more basic than data and that concepts determine the employment of indicators. The old problem of striking a balance between the theoretical and empirical elements of social science knowledge is thus solved by *theoreticism* arguing that the choice is between alternative theories according to theoretical criteria. This starts the vicious circle that theories cannot be compared on the basis of their claims to truth but the choice of a theory is simply a function of the perspective adopted, which in turn follows from the theory. Thus we move in a closed system: the pros and cons of various approaches and perspectives cannot be resolved by evaluating the truth claims of various theories by means of a neutral body of data. It is argued that the concepts employed in comparative inquiry cannot be related to a common reality by means of a set of indicators, because the various concepts have their own indicators.

The dogmas of theoreticism are as exaggerated as the crude versions of logical empiricism. Both fail to strike a balance between theory and data, concepts and indicators. However, social science inquiry needs both, without any dogmas as to which is basic: theory without data is empty and data without theory is blind. The distinction between theory and data, concepts and indicators may be stated without a double language theory. Moreover, it is possible to show that empirical considerations are relevant to theory formation and concept construction (Bollen, 1980; Marradi, 1981).

To the social analyst the indicator approach has a distinctive advantage compared with the alternatives of empiricism or theoreticism, because the indicator approach accomplishes a balance between theory and data. It sustains the fundamental idea that theories may be wrong due to the outcomes of the application of the indicators. However, it could be argued that the indicator approach may be interpreted in terms of theoreticism (indicators as conventions) or in terms of empiricism (indicators reducing constructs to observation).

Indicators as conventions: the convention interpretation of the concept of an indicator implies that indicators may be specified arbitrarily. It is a completely open choice to be compared with the stipulation of terms. The restrictions are theoretical, deriving from the consequences of the convention for other indicators. Thus, what matters is the theoretical context into which the indicator and its construct enters – theoreticism. The basic counter-argument is that the choice of indicators has consequences not only for the meaning postulates of a theory or to put if differently, the choice of meaning postulates (indicator-construct relationships) cannot be made without restrictions deriving from the data.

Indicators are not simply observation or observation terms; they are different in nature from the elements of an empiricist observation language. Indicators require for their construction the sort of imaginativeness and hindsight that are typically expressed in theory construction. Indicators or indices are tools for the description of data; they are not in themselves observation statements, as such propositions were conceived by logical positivists. Denying that there is such a thing as a theory neutral

or theory independent observation language does not imply that there is no way to report on data that can be used as a test of a theory.

We have to make a distinction between the theory T that relates the constructs to each other and the theory that introduces the indicators that allow us to report on observations and speak about the data $-T^*$. And it is not the case that $T = T^*$ or that T implies T^*. By creating relationships between T or a set of T:s ($T_1, \ldots T_N$) and T^* it becomes possible to test the propositions of T and to measure its variables. Such relationships may be identified as correspondence rules, meaning postulates or operational definitions in order to identify such relationships as something different from statements of interaction between theoretical variables or statements of correlations between indicators. But the fact that such measurement rules in terms of which a construct is tied to an indicator or index cannot be said to be true or false does not imply that they are purely conventions. Conventions are arbitrary, just like naming or definitions of terms; rules governing the construct-indicator relation are appropriate or nonappropriate, a matter to be judged in terms of criteria of validity and reliability.

Indicators as causes or effects: the idea that indicators are the manifest expressions of their constructs implies an ontological commitment. It implies either that constructs may interact with measures or scores or that there is a world of latent phenomena that exists though it is not visible except by means of the indicators. This is not the place to enter the debate between realists and instrumentalists about what exists (Nagel, 1961), but it seems possible to treat the question of the relationship between theory and data as separate from the general problem of the nature of theoretical entities (the cognitive status of theories). Whether indicators are conceived of as causes or effects of constructs or latent variables or not it is still true that indicators are necessary for measuring constructs, and without measurement there can be no test of theoretical propositions.

8. Conclusion

Following the Kant argument it is necessary to strike a balance between theory and data. However, the solution to the problem of identifying how theory and data may interact and yet not be reducible to each other cannot be found in the traditional distinction between construct and observation as Kant conceived it:

"Our knowledge springs from two main sources in the mind, the first of which is the faculty or power of receiving representations (receptivity for impressions); the second is the power of cognizing (spontaneity in the production of conceptions). Through the first an object is given to us; through the second, it is, in relation to the representation (which is a mere representation of mind), thought." (Kant, *Critique of Pure Reason*, 1978: 62)

By framing the problem of the theory-data interaction in terms of the distinction between concept and observation one is forced to move towards a choice between either empiricism or theoreticism. If theories are to be evaluated in terms of data, and if observation determines a theory neutral set of data, then alternative theories may be tested against a common ground; but if there are no theory neutral observations, then how can theories be tested?

The construct-indicator distinction appears to be more congenial to the theory-data interaction problem, because it does not require notions about an empiricist language or a double language model. It sustains the fundamental idea that theories may predict different propositions about the data, which may be true or false. The concept of an indicator has its own problems – what is the character of the relationship between an indicator or several indicators and the construct it (they) indicates (indicate)? Yet, an indicator is not an observational term like the logical positivist had in mind and it allows for the measurement of the construct without entering into the theory in which the construct is placed. The application of a preliminary set of indicators to data may have theoretical implication for the further elaboration of constructs. Thus, far from each theory determining its own indicators, alternative theories may be tested against a set of indicators – a test that carries implicators for the theoretical elaboration of constructs. Without the indicator the construct cannot be measured, and thus it carries no implications for the data. Without the construct the indicator lacks an interpretation spelling out the meaning of the data and what is measured. Construct without indicator is empty and indicator without construct is blind.

References

Blalock, H.M. (Ed.): Causal Models in the Social Sciences. Chicago: Aldine Publ. Co. 1971.

Blalock, H.M.: Social Statistics. New York: McGraw-Hill 1972.

Blalock, (Ed.): Measurement in the Social Sciences. Chicago: Aldine Publ. 1974.

Bollen, K.A.: Issues in the Comparative Measurement of Political Democracy. In: American Sociological Review, 45 (1980), p. 370–390.

Feyerabend, P.: Against Method. London: Verso 1975.

Kant, I.: Critique of Pure Reason. New York: Macmillan 1978.

Kuhn, T.S.: The Structure of Scientific Revolutions. Chicago: The University of Chicago Press 1962.

Lazarsfeld, P., Rosenberg, M. (Eds.): The Language of Social Research. New York: The Free Press 1955.

Lazarsfeld, P., Pasanella, A.K., Rosenberg, M. (Eds.): Continuities in the Language of Social Research. New York: The Free Press 1972.

Lijphart, A.: Comparative Politics and the Comparative Method. In: American Political Science Review. 65 (1971), p. 682–693.

Marradi, A.: Factor Analysis as an Aid in the Formation and Refinement of Empirically Useful Concepts. In: Jackson, D.J. & Borgatta, E.F. (Eds.): Factor Analysis and Measurement in Sociological Research. London: SAGE 1981, p. 11–49.

Morgan, G.: Beyond method. Strategies for social research. London: SAGE 1983.

Northrop, F.S.C.: The Logic of the Sciences and the Humanities. New York: Macmillan 1947.

Przeworski, A., Teune, H.: The Logic of Comparative Social Inquiry. New York: Wiley 1970.

Sartori, G. (Ed.): Social Science Concepts. Beverly Hills: SAGE 1984.

Sartori, G., Riggs, F.W., Teune, H.: Tower of the Babel: On the Definition and Analysis of Concepts in the Social Sciences. Pittburgh: International Studies Association 1975.

Sjöblom, G.: The Cumulation Problem in Political Science: An Essay on Research Strategies. European Journal of Political Research, 5 (1976), p. 1–32.

Sullivan, J.L., Feldman, S.: Multiple Indicators. Beverly Hills: SAGE 1979.

Prof. Dr. J.-E. Lane, Department of Political Science,
University of Umea, S-90187 Umea, Sweden