THE TRANSFORMATION OF COGNITIVE VALUES INTO METHODOLOGICAL RULES

Erik Weber

1. *Introduction*

Methodological research in the strict sense - the construction of rules of acceptance and rejection for scientific theories - was one of the main topics of philosophy of science during the late sixties and a large part of the seventies (the period of Popper, Lakatos, Feyerabend and Laudan's Progress and its Problems). But during the last decade, most energy has been spent to other problems: incommensurability and underdetermination, the possibility of a logic of discovery etc. Another "new" topic was the analysis of scientific explanations and other aims of science ("cognitive values"; note that this problemshift is mirrored in the work of Laudan: "Science and values" deals with cognitive values, while "Progress..." still dealt with acceptance and justification).

The conclusion is obvious: the least we can say is that this paper is a little bit out of time. So why have I written it? My aim is twofold. First of all, I'd like to show that the results of the studies of the aims of science allow us to give a new impulse to methodological research. I mean that, by taking these results into account, we may avoid coming to a dead end, like Popper, Lakatos etc. Taking the studies of cognitive values seriously implies that two fundamental questions are to be raised: (1) Are methodological rules useful everywhere? Or is their value connected with some cognitive value (which would mean that they're only useful in some limited contexts). (2) Which conditions does the study of cognitive values impose on the construction of methodologies? What are the basic characteristics that are entailed by certain principles of the analysis of the aims of science? Answering these questions is my first aim. My second (but not less important) aim is to develop a set of methodological rules which conforms to the answer that will be given to the questions above.

Section 3 of this paper consists of a short analysis of the aims of science, based on the most important ideas of the relevant literature and on some personal elements. This analysis has a double function. Firstly, it is the starting-point for the deduction of methodological rules (the method of deduction is explained in section 2, the deduction itself is the subject of section 4). The other function of this analysis is to serve as a basis for reflections on our fundamental questions (our first aim). Because these reflections are dealt with in the fifth and last section, the methodology presented in section 4 is uncritical and therefore provisional: large parts of it will have to be removed when we've finally answered the questions above. I don't think this is a waist of time. I'm convinced that it is the clearest possible way to present my ideas.

2. *Preliminary remarks: how to develop a methodology?*

The function of a set of methodological rules is to predict to which extent the current scientific theories allow us to reach the aims of science. Since I assume that this aim is threefold (explanation, control, prediction; cfr. section 3), one could say that a methodology must enable us to predict whether we will be successful in explanation, prediction and/or control if we apply the theory that is being evaluated. Predictive power is an essential feature of methodologies: they try to predict the success of a theory *before* this theory is applied. Of course these predictions must be reliable, in other words: the methodology must be epistemologically relevant. The procedure I propose to guarantee both predictive power and epistemological relevance, is the following:

1. Describe the different aims of science (i.e. formulate criteria of adequacy for each of them, as will be done in section 3).
2. Try to transform the criteria on explanations etc., obtained in the first step, into requirements about the features of a *theory*. If this transformation is impossible for one of the criteria, it is to be neglected.
3. Find an operational procedure to test the presence of each of the characteristics obtained in the second step.

ad 1: Starting from the analysis of the aims of science guarantees that our rules are epistemologically relevant. The only problem is that not each criterion can be transformed, which will cause a degree of uncertainty (see *ad 2*).
ad 2: The characteristics of "good" explanations etc. are formulated as features of theories: this way we obtain the characteristics of "good" theories. This second step guarantees the pre-

dictive power, but from now on some aspects of the aims of science (the criteria referring to singular statements and those referring to pragmatical circumstances) are neglected, because transformation is impossible. Each rule we obtain is relevant, but we don't have enough rules to eliminate false predictions completely.

ad 3: The third step will be trivial in some cases: some of the features described will be immediately operational. Other features will require partial application of the theory to a limited group of problems. The information thus obtained is extrapolated, inductive steps are taken. This is a second source of uncertainty.

This procedure will be used in the following sections. The first step is executed in section 3, the other steps in section 4.

3. *A starting-point: the aims of science*

3.1 This survey will be as short as possible. Within the scope of this paper, it is impossible to give any arguments for what I will say in this section. I'm sure that the reader will be able to distinguish traditional and personal elements, so I won't separate them clearly. On the other hand, I hope that those who tend to reject my analysis at first sight, will be convinced by the arguments developed in the literature it is based on.[1] One can distinguish three aims of science: prediction, control and explanation. These aims can be realized by producing arguments of a certain kind: predictive arguments, "control-arguments" and explanatory arguments. An analysis of the aims of science must consist of a complete list of criteria to be fulfilled by these arguments in order to be adequate. Such a list will be given for each kind of argument.

3.2 *Predictive arguments*

The structure of a predictive argument can be represented as follows:

$$\frac{\begin{array}{l} C_1, C_2, \ldots, C_n \\ L_1, L_2, \ldots, L_n \end{array}}{P}$$

where P and $C_1 \ldots C_n$ are singular statements, $L_1 \ldots L_n$ law statements, and the single line indicates a derivability relation (deductive or inductive). The first criterion that is to be introduced determines the pragmatic circumstances:

(P1) The truth-value of P was not known before the predictive argument was given: our knowledge about P depends on the argument.

This means that the initial conditions and law statements are not only the logical starting-point, but the pragmatical starting-point too. Two criteria are to be introduced as minimal requirements for potential predictions:

(P2) The antecedent part of a predictive argument contains at least one singular statement.

(P3) The antecedent part of a predictive argument contains law statements which connect the events described in the initial conditions and those described in the conclusion in the following way: $P(p \mid c_1 \ldots c_n, b) = r$, where $P(p \mid c_1 \ldots c_n, b) \neq P(p \mid b)$. (*b* are the circumstances)

These two criteria define exactly what is represented in the scheme above. Since scientists want to be able to predict correctly, an additional criterion must be found to discern "correct" predictions and merely potential ones. What is a "correct" prediction? Obviously, the conclusion of the prediction is decisive here: a correct prediction is one with a correct conclusion. The conclusion can be said to be correct if it may be confirmed by means of empirical evidence (it is assumed that empirical evidence is only obtained after the prediction has been uttered, see (P1)). This requirement of convergence empirical and rational evidence may be formulated as follows:

(P4) When empirical evidence about the truth-value of the conclusion is obtained, this evidence must confirm the result of the predictive argument.

(P4) can be seen as combining two requirements about the *premises* of the argument (when we assume that the convergence is not due to pure chance):

(P4a) We must have empirical evidence which confirms the initial conditions.

(P4b) The alleged statistical relevance relation (cfr P3) must be tested.[2]

(P1) till (P4) are the minimal criteria to be imposed on predictive arguments. The subsequent criteria give rules for making a selection between alternative (i.e. competing) minimally sufficient predictions:

(P5) Determinism (in a positive or negative sense: the antecedent part deductively entails P *or* not-P) is an ideal.

(P6) If all the competing predictions are statistical, Hempel's "Requirement of maximal definiteness" is to be followed.

3.3 *"Control-arguments"*

Our starting-point is the same as for predictive arguments, viz. the following scheme:

$C_1, C_2, ..., C_n$
$L_1, L_2, ..., L_n$

G

G describes the event we want bo bring about (the "goal"). The pragmatic circumstances can be formulated as follows:

(C1) The truth-value of G is fixed and known: G is false and we know it. Our aim is to change the truth-value of G (i.e. to bring about the event described by it) by realising certain initial conditions previously unfulfilled.

There are two minimal requirements for a potential control-argument:

(C2) analogue to (P2)

(C3) The antecendent part of a control-argument must contain law statements which jointly assert that $c_1... c_n$ is a cause of g (where $c_1...c_n$ and g are the events described by $C_1...C_n$ and G) in circumstances b.

Actually, this latter claim combines two requirements:

(C3a) The argument must contain law statements which jointly assert that $P(g|c_1...c_n,b)=r$, with the restriction that there must be a positive SR-relation: $P(g|b,c_1...c_n) > P(g|b)$.

(C3b) It must be asserted that there is a productive power of $c_1...c_n$ with respect to g: we can produce g by means of $c_1, ...c_n$, or at least have a positive influence on its realisation.

Of course, what we are interested in are only those control-arguments that really are effective. The analogon of a correct prediction can be defined by adding (C4) to the previous criteria:

(C4) A control-argument is really effective if and only if: (i) we are able to realise the events described in its initial conditions (either by performing a simple action or by means of another control-argument), and (ii) the strategies described in the law statements prove to be effective (since we don't have a method to anticipate this, the effectiveness can only be verified post factum[3]).

Again we have two criteria which give rules for choosing between rivalising control-arguments:

(C5) Determinism (in a positive sense: only deduction of G) is an ideal.

(C6) If all the competing arguments are statistical, the one with the highest a posteriori probability of G must be chosen.

3.4 *Explanatory arguments*

Again our starting-point is the structure

$C_1, C_2, \ldots, C_n$
$L_1, L_2, \ldots, L_n$

E

E describes the event to be explained. The pragmatic circumstances are as follows:

(E1) The truth-value of E is fixed and known: E is true. The problem is that this is surprising: though we know that E is true, we don't understand why.

Two minimal requirements define the concept of a potential explanation:

(E2) Analogue to (P2) and (C2)

(E3) The explanans must contain law statements whose conjunction asserts that $c_1 \ldots c_n$ is a cause of e (=the explanandum event) or of not-e.

Like (C3), (E3) is a combination of two partial criteria, which are analogue to (C3a) and (C3b).

Explanations which meet the criteria (E1)-(E3) are minimally sufficient: no additional criterion is required. This means that the components of the explanation don't need verification by means of some testing procedure: it is sufficient that we *assert* that here is a causal connection. There are several criteria which are to be combined in order to select an ideal explanation:

(E4) Determinism (only in the positive sense) and *approaching* determinism (in both senses) are ideals.

(E5) Continuity criterion: an explanation is better if there is more continuity between the successive stages of the causal mechanism as it is described in the causal laws of the explanans.

(E6) Fundamental mechanism criterion: the explanatory power increases when the causal mechanism is analysed in terms of the mechanisms that, in a certain discipline, are seen as intuitively clear and not in need of further analysis.[4]

Although they define ideal explanations, (E5) and (E6) are, to some extent, minimal requirements too: a minimal degree of continuity is to be guaranteed to make an explanation adequate; when the fundamental mechanisms are totally neglected, the explanatory value is nihil.

(E7) An explanation is to be preferred if parts of its explanans are identical to or can be derived from the explanans of other explanations.

(E8) An explanation is to be preferred if its law statements and initial conditions are maximally consistent with the explanans-part of other explanations.[5]

(E7) and (E8) both refer to the ideal of explanatory unification. (E7) is the strongest criterion: it entails (E8), but not vice versa.

4. *Methodological rules*

4.1 The analysis of section 3 has revealed that there are important structural differences between explanations, predictions and control-arguments. Not each explanation can be transformed (by adjusting the pragmatic circumstances) into a prediction, nor vice versa. The same may be said about explanations and control-arguments and about predictions and control-arguments. The three aims of science are not convergent. The same theory may e.g. do very well at explaining, and may be quite useless for prediction. Therefore it is impossible to construct a universal set of methodological rules which would cover the three areas. Instead one has to develop three groups of rules for acceptance, to be used alternatively, depending on the kind of application we have in mind for the theory.

In the previous section we distinguished minimum criteria on explanation, control and prediction from the definition of ideals. The methodological rules derived from the different sorts of criteria will have a different status. A set of minimal criteria may be transformed into a group of rules which determine whether a theory has any scientific value at all: these rules contain absolutely necessary minimal conditions. Consequently, this first step of the evaluation is *not comparative*: competing theories are irrelevant. Sets of requirements of the second kind may be transformed into *comparative* rules to select the best theory among those who passed the first tests.

The two remarks above explain the structure of this section: for the sake of simplicity I shall deal with the non-comparative rules first (4.2), and discuss the comparative rules later (4.3). Because of the structural differences mentioned above, each part will consist of three sets of rules.

4.2 *Non-comparative methodological rules*

Explanation

We start from the criteria (E1)-(E3). If we execute the second step of our production scheme, (E1) and (E2) are eliminated. On the other hand, restating (E3) yields the following result:

(T1) A theory must contain statements asserting the existence of (probabilistic or deterministic) causal connections between types of events: the occurrence of an event of type A (always or in X% instances) causes the occurrence of an event of type B, in circumstances C. (A and C are usually compound)

What is the operational test to decide whether a theory conforms to (T1)? It is obvious that no specific tool or procedure is necessary to test here: (T1) is operational of itself. Consequently, our first methodological rule is simply:

(MR1) Verify whether (T1) is fulfilled.

Control

We start from the criteria (C1)-(C4). (C1) and (C2) are to be neglected, since they don't refer to theories. If we transform (C3), the result is (T1). (C4) can be transformed partially, which gives us (T2):

(T2) The alleged causal laws must describe strategies that really are effective.

This last requirement may be divided into (T2a) and (T2b), which respectively deal with the statistical-relevance-aspect and the productivity aspect of causation:

(T2a) It must be proved that there really is a positive statistical relevance relation $P(B|A_1...A_n,C)>P(B|C)$.

(T2b) It must be proved that there really is a productive power of A with respect to B.

Unlike (T1), (T2) is not operational in itself. Since the effectiveness of a strategy has two aspects, we can develop two tests, each of them accounting for one element. Regarding (T2a), I claim that a statistical relevance relation can be tested by comparing the results of two experiments, which resp. determine the relative frequency of B's in samples of the reference class $A_1\&...\&A_n\&C$ and of the reference class C. The ratios of the two samples are supposed to be exact indicators of the ratios in the whole reference class. This means that I hold on to Reichenbach's straight rule of induction: this rule is to be applied twice and the resulting probabilities must be compared to verify whether they are different:

(MR2a) In order to test $P(B|A_1...A_n,C)>P(B|C)$,

1. determine the relative frequency of B in a sample of both reference classes,
2. extrapolate to the total reference class (i.e. adjust the initial hypotheses about the relative frequencies in function of the results of 1),
3. verify whether the former probability in 2 is greater than the latter.

There is no theoretical justification for the assumption that Reichenbach's straight rule is valid. The only way in which (MR2a) can be justified is by means of a pragmatic second order induction: in the past this methodological rule has been more successful than its competitors in predicting success with regard to the aims of science (in this case: is guaranteeing successful manipulation and intervention). The rule which tests the existence of a productive power in the appropriate direction completes the non-comparative part with respect to control. The intuitive solution is the pragmatical one: given the fact that there is a positive SR-relation, a sufficiently great sample is chosen to test the direction of the productive power in practice, in the following way:

(MR2b) A ($=A_1...A_n$) is causally prior to B if and only if: if the relative frequency B/A equals r (in circumstances C), we can produce/bring about a B by doing A in r instances of the sample.

This rule can be justified in a way similar to (MR2a). It should be noticed that (M3b) is not applicable in areas where human intervention is impossible. But since the range of control-arguments is obviously limited to the domain where human intervention *is* possible, there is no problem: the lack of a criterion that exceeds the area of possible human action is harmless because beyond this area we can only try to explain or to predict, which means that (T2b) is irrelevant.

Prediction

We start from the criteria (P1)-(P4b). (P1), (P2) and (P4) are to be neglected. If a theory is to be used for prediction, the following characteristics are relevant at the first stage of its evaluation:

(T3) A theory must contain statements asserting that there is a (positive or negative) statistical relevance relation between events of type A and type B: $P(B|A\&C)\neq P(B|C)$, where C are the circumstances.

(T4) The SR-relation referred to in (T3) must be verified.

(T3) and (T4) were derived from resp. (P3) and (P4b). The

methodological rule corresponding to (T3) is simply:
(MR3) Verify whether (T3) is fulfilled.
In order to test (T4), it is sufficient to adjust (MR2a) so as to make it refer to negative statistical relevance too.

4.3 *Comparative methodological rules*

If a theory passes the relevant tests among (MR1)-(MR3), we can be sure that it is a useful theory, that it will guarantee success in explanation, prediction or control to some degree. In order to maximize this success, competing theories must be compared in relation to the characteristics that can be derived from the requirements defining ideal explanations, predictions and control-arguments.

Prediction
From (P5) and (P6) we can derive two features that are important for the further evaluation of theories to be used for predictions:
(T5) Deterministic theories are to be preferred.
(T6) If none of the competing theories is deterministic, the definiteness of the reference classes must be taken into account: the higher this definiteness, the better the theory.
The relevance of determinism and maximal definiteness is due to the ideal of complete information: only deterministic deduction assures that the information was complete, while maximizing the definiteness of the reference class means that all known information is taken into account.

Control
Again we have to relevant characteristics: (T5) and (T7):
(T7) If none of the competing theories is deterministic, the one which assigns the highest a posteriori probabilities must be chosen.
The justification is simple: all that matters here is a greater chance to obtain a goal; so the a posteriori probability is to be maximized (cf. C5 and C6).

Explanation
With respect to explanations, we can derive at least six criteria for further evaluation. Because these criteria (unlike those for prediction and control) may be contradictory, a hierarchy (the great lines of which are already given by the place of each criterion in the list below) will be proposed.

(T8) Theories which describe causal mechanisms in a continuous way are to be preferred.

(T9) Theories that make use of the "fundamental mechanisms" are to be preferred.

(T10) Theories with greater empirical content are to be preferred.

(T11) Pairs or groups of theories which show certain similarities or analogies, must be preferred.

(T12) Contradictions between theories are to be avoided as much as possible.

(T5) Deterministic theories are to be preferred.

Three categories can be distinguished in our list: (T8) and (T9) deal with ontological aspects (cfr. (E5) and (E6)). (T10)-(T12) refer to explanatory unification. (T10) covers the internal aspect of unification (viz. explaining as much as possible by the same premisses, i.e. by the same theory), while (T11) and (T12) cover the external aspect (congruence of explanatory theories). Finally, (T5) deals with determinism. Because the ontological aspects described in (T8) and (T9) are, in my opinion, more decisive for the explanatory power of an argument, (one must remember that (E5) and (E6) to some extent were conditions for minimally sufficient explanations) than e.g. the aspect of explanatory unification, I am convinced that the first two criteria of our list are to be placed at the top of the hierarchy. Since they aren't contradictory, no further specification is necessary. The criteria covering the aspect of explanatory unification form the second level. The internal hierarchy of this level is clear: one unified theory is better than mere similarities between separate theories. Finally, determinism is at the lowest level because one may wonder whether this ideal is attainable at all. Isn't there a gap between the ideal of deductive explanation and the basic structure of the world (which, after all, may turn out to be probabilistic?)

4.4 *Summary*

To conclude this section, it might be useful to present the results we obtained in a schedule (the next page).

As most readers presumably already have noticed, I didn't mention the operational counterparts of (T5)-(T12). To my view there are no special problems, at least when concepts like continuity and empirical content are properly defined. This means that the methodological rules can be obtained in the same trivial way as we used for (T1) and (T3).

	EXPLANATION	CONTROL	PREDICTION
NOT COMPARATIVE	T1 (MR1)	T1,T2a,T2b (MR1,MR2a, MR2b)	T3,T4 (MR3,MR2a)
COMPARATIVE	- T8,T9 - T10,T11,T12 - T5	T5,T7	T5,T6

5. *General conclusions*

What does the previous section, i.e. the attempt to derive a set of methodological rules from a description of the aims of science, tell us about the fundamental questions raised in the introduction of this paper? Can we decide now within which limits methodological research is possible and/or useful? Are we able to decide on conditions that, even within these limits, are to be respected by any attempt to develop methodological rules? As we will see in this section, the answer to the last two questions must be affirmative.

If we look back at section 4 and wonder whether we have described a useful methodology or not, the answer is twofold: some parts are useful, others certainly are not. The fundamental problem is the following: the methodology I proposed, is - like all methodologies - holistic and abstract, which means that it evaluates theories without referring to the problems these theories are supposed to solve. Only the *kind* of problem (predictive, explanatory, ...) is considered to be relevant. This is an inescapable consequence of the function of methodologies. A basic question then is: how can we justify this holistic and abstract approach? What's its relevance? Why not wait till we've really been confronted with some problem, and then choose the best theory for it, independently of the solution of other problems? As a matter of fact, I'm convinced that this casuistic approach is the right one in the areas of prediction and control. In other words: it is useless to evaluate and compare theories with regard to their overall predictive power or their overall efficiency. The decision to prefer one theory or another must be taken over and over again, each time a new problem of prediction or human intervention occurs. Why? The analysis of control-arguments and predictions didn't reveal any need for unification in these areas: the adequateness of a prediction or control-argument is totally independent of the way we solve other predictive or productive problems. Since it is therefore very unlikely that one theory will be best suited in all possible cases, methodological rules are without any use here. Moreover, it must be noticed that the application of (T6) and (T7) would have caused considerable metric problems: theories make assertions about several reference classes, and assign lots of a posteriori probabilities. How would we have compared this?

When we look at the area of explanation, a different answer must be given: methodological rules are useful here because of the ideal of *explanatory unification*. Because the adequateness of an explanation partly depends on the way in which we solve other explanatory problems, the holistic approach makes sense in this context. We confine the task of methodological evaluation to the determination of the explanatory power of a theory. Consequently, the analysis of scientific explanations is the only starting-point for methodological research. Large parts of section 4, viz. those dealing with prediction and control, are to be neglected (at least as an attempt to develop methodological rules: (MR2a) and (MR2b) are for instance indispensable if we want to make the criteria on prediction and control operational; so they have their function in the casuistic approach).

Granted methodology is to be confined to the area of explanation, a further question is whether our analysis has revealed some general conditions to be met by further research. Without assuming that my proposal is completely adequate, one may conclude that most of Hempel's heritage is to be neglected: explanations are more than derivations. Explanations are requests for causes, we have to place the problematic event in a causal network (cfr. W. Salmon, 1984). Such a causal approach of explanations combines three aspects: (i) the derivational aspect (cfr. Hempel), (ii) the aspect of productivity/efficiency (cfr. (E3)), and (iii) the ontological aspect (cfr. (E5), (E6)). Only this causal approach, combined with the idea of explanatory unification, can lead to an adequate methodology. The paragraphs of 4.2 and 4.3 dealing with explanations are an example of such a methodology. It is obvious that other proposals, based on a different analysis of explanation and causation, may be developed. This once more illustrates the importance of these latter analyses: they will be indispensable if we have to choose between the different proposals.

NOTES

1. A complete survey of this literature is impossible, but the list

of references contains the most important books.
2. An operational test will be described in 4.2.
3. see (2)
4. Examples of these fundamental mechanisms are: stimulus-response-scheme in psychology and processes of attraction and repulsion at atomic level in physics.
5. The inconsistencies meant here are not only those of classical logic. For instance: A causes B is inconsistent with B causes A, because of the asymmetry of causal relations, but this goes beyond classical logic.

REFERENCES

Gärdenfors, P., (1980), "A pragmatic approach to explanations", in *Philosophy of Science*, 37, pp. 405-423.

Hempel, C.G., (1965), *Aspects of scientific explanation and other essays in the philosophy of science*, New York, Free Press.

Lakatos, I., (1970), "Falsification and the methodology of scientific research programmes", in *Criticism and the growth of knowledge*, edited by I. Lakatos and A. Musgrave, London, Cambridge University Press, pp. 91-95.

Laudan, L. (1977), *Progress and its problems*, Berkeley, University of California Press.

Laudan, L. (1984), *Science and values*, Berkeley-Los Angeles-London, University of California Press.

Popper, K.R., (1959), *The logic of scientific discovery*, New York, Basic Books.

Popper, K.R., (1972), *Objective knowledge*, London, Oxford University Press.

Reichenbach H., (1938), *Experience and prediction*, Chicago, University of Chicago Press.

Salmon, W., (1984), *Scientific explanation and the causal structure of the world*, Princeton, New Jersey, Princeton University Press.

Stegmüller, W., (1983), *Probleme und Resultate der Wissenschaftstheorie I. Erklärung, Begründung, Kausalität*, Berlin, Springer Verlag (2nd, improved and enlarged edition).

Suppes, P., (1970), *A probabilistic theory of causality*, Amsterdam, North Holland Publishing Company.

Van Fraassen, B., (1980), *The scientific image*, Oxford, Clarendon Press.

Von Wright, G.H., (1971), *Explanation and understanding*, Ithaca, New York, Cornell University Press.