First published in *Towards Unification in Psychology*, J. R. Royce (ed.), 1970. Extract, pp. 54–103. Toronto: University of Toronto Press.

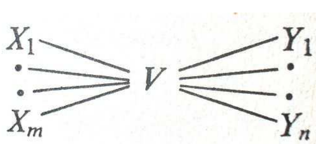
The Art of Metascience; or, What Should a Psychological Theory Be?

Introduction

In order not to lay false claims upon your attention, let me confess at once that the subtitle of this paper is intended more as a stimulus to curiosity than as a serious synopsis of the subject here to be addressed. I shall indeed be concerned at considerable length with the methodological character of psychological theories, but my prescriptions will focus upon process rather than product. Specifically, I shall offer some detailed judgments on how theorizing must be done—or, more precisely, what sorts of metatheorizing must accompany it—if the enterprise is to make a genuine contribution to science; and while on first impression these may seem like no more than the standard broad-spectrum abstractions of one who has nothing specific to say, I assure you that they envision realizable operations which, if practiced, would have a profound effect on our conceptual efficiency as psychologists.

There are at least two formal dimensions along which discussions of psychological theory can vary. One is *level of abstraction*, or *substantivity of concern*, with attention to specific extant theories (e.g. pointing out an inconsistency or clarifying an ambiguity in John Smiths' theory of binocular psychokinesis) at one extreme, and philosophy-of-science type concerns for the nature and functioning of idealized theories in general, detached from any real-life instances, at the other. Distinct from this, though not wholly orthogonal to it, is a second dimension of metatheory which might be called *acuity* or *penetration*, and which concerns the degree to which the discussion makes a serious, intellectually responsible attempt to further our understanding of the matter with which it deals. Here the possibilities range from painstaking attention to technical details, to loose and largely gratuitous generalities built around everyday intuitions, or poorly defined neologisms whose literal relevance to anything in reality is tenuous or nonexistent. Unfortunately, disciplined thinking *about* science appears to be much more difficult to achieve than disciplined thinking within it, with the result that to date, very little metascience has managed to get far from the casual, dilettante or intuitional end of the acuity scale, especially at the higher levels of abstraction. Since metatheoretical acuity is a major concern of this essay, let me illustrate what I mean by this with three examples of inadequate penetration in abstract metatheory. The first two are narrow aspects of a recently notorious but now moribund metapsychological issue which today can be cited briefly and dispassionately even while a sense of its historical vitality still lingers. The third has contemporary philosophical immediacy, but as yet little if any impact on psychology proper. All three lie on the respectable side of the acuity distribution—at the other end, ideas are muddled about so inchoately that one can scarcely catch hold of anything firm enough to criticize.

(i) Some time ago, when the nature of "intervening variables" was a major bone of metapsychological contention, Feigl (1945) suggested that the conceptual virtue of a single intervening variable V interposed between m independent variables $X_1, \dots X_m$ and n dependent variables $Y_1, \dots Y_n$ might be that the number of laws relating a dependent variable Y_i to an independent variable X_j becomes reduced from the $m \times n$ possibilities $Y = \phi_{ij} X_j (i = 1, \dots, n;\ j = 1, \dots, m)$ to the more parsimonious $m + n$ array $V = \mathrm{g}_j(X_j), Y_i = \mathrm{f}_i(V)(i = 1, \dots, n; j = 1, \dots, m)$, wherein each observable relation ϕ_{ij} is analysed as the product $\mathrm{f}_i \mathrm{g}_j$ i.e. $\phi_{ij}(X_j) \equiv \mathrm{f}_i(g_j(X_j))$. This model, often expressed by a visual diagram of form



was quickly seized upon as the standard conceptualization of intervening variables, even to the point of serious interpretations of empirical findings in its terms (e.g. e.g Seward, 1955; Miller, 1959, p. 276f.), with no apparent concern for what could possibly be *meant* by the pairwise relationships ϕ_{ij} and the visual model showing V at the focus of lines converging from the various independent variables. Dependent variable Y_i (or V) does not have a separate deterministic relation to each of the m independent variables. There is only a single equation $Y_i = \phi_i(X_1, \dots, X_m)$ determining Y_i as a joint function of the variables $X_1, \dots, X_m$ which affect it. Thus introduction of an intervening variable between the X_j and the Y_i replaces the n equations $Y_1 = \phi_1(X_1, \dots, X_m), \dots, Y_n = \phi_n(X_1, \dots, X_m)$ with the $n + 1$ equations $V = \mathrm{g}(X_1, \dots, X_m), Y_1 = \mathrm{f}_1(V), \dots, Y_n = \mathrm{f}_n(V)$—whence the argument from parsimony must always protest against introduction of the intervening variable. This is not to suggest that no interpretation can be found for the pairwise-relations model—I can think of several, though each has its own methodological complications.[1] The point is that no such interpretation was, in fact, ever provided for it in the literature. A good fifteen years of metatheorizing about the role of constructs

[1]The best is a Spearman single-factor model in which relation-term ϕ_{ij} is interpreted as the linear correlation between variables Y_i and X_j. However, ϕ_{ij} is then no longer the functional dependence of Y_i upon X_j but only a measure of how errorlessly linear the relationship is.

in psychological theory allowed a primary sector of its thinking to rest upon a simple fallacy which could have been rehabilitated into fruitful liaison with inferential factor theory by a few moments of critical contemplation.

(ii) Another notion which cropped up repeatedly in the intervening variable literature was that when an intervening variable V is introduced into the relationship $Y = \phi X$ between data variables X and Y, V is *the relationship itself.* But this is flagrantly inconsistent with what were cited as paradigm cases, notably, behaviour-theoretical variables such as Hull's Habit-strength, in which the intervening variable V is functionally dependent upon X and in turn determines Y. Regardless of whether V mediates causally between X and Y or is merely a logical construction, it is still a variable which *partakes* of relations to the data variables, in conspicuous logical contradistinction to the fixed function ϕ in $Y = \phi X$ which is not *related* to the data variables but is the relationship itself. What is especially ironic about this confusion of variables which intervene with the relationship *within* which they intervene is that at a higher level of logical complexity, data-variable relationships which are themselves variable turn out to be a primary source of inferences to underlying states of the organism which, moreover, are importantly distinct from the internal processes which mediate the relationships from which these states are inferred.[2] Before the logical intricacies of this situation can be effectively navigated, however, it is essential that the concept of "variables" and their relationships be understood with a modicum of technical precision.

(iii) In rebellion against the "covering law" interpretation of scientific explanation, a view has recently arisen in philosophical quarters that the breath of life in an explanatory theory is a distinction between what is "natural" and what is not, namely, that the theory makes first and foremost a commitment to an "ideal of natural order" which is self-explanatory simply because it *is* "natural," and then seeks to account only for apparent departures from this ideal (Toulmin, 1961). The paradigm example for this view is the shift from Aristotelian to Newtonian conceptions of natural motion, the former having expected a moving body to come to rest if left to itself while the latter expects the motion of an unconstrained body to persist without change. There is certainly an intuitive plausibility to this thesis, yet a critical search for the role of "naturalness" in the actual conceptual impact of a theory such as Newton's laws of motion has much the same success as looking inside one's television set for the little people who put on the show. The theory simply states what happens (or what *must* happen, if we wish to emphasize the theory's nomic character) under different configurations of values for the relevant variables, and "natural ideals" enter only in the sense that some configurations have simpler consequences than others. Thus Newton's laws tell how a

[2] The distinction between state and process variables, and inference to underlying states from mutable parameters in observed process regularities, are discussed at somewhat greater length in Rozeboom (1965)

body moves (or what forces direct its motion) for any arrangement of surrounding bodies, and the empty-surround condition is simply one limiting possibility which has no theoretical priority over any other. Neither does the Newtonian theory explain only departures from the "natural ideal" of unconstrained motion. A system's temporal progress from any one initial configuration bears the same logical relation to the theory's postulates as does any other, and the theory "explains" motion in the absence of constraints fully as much and in the very same way that motion in more complex circumstances is explained. I do not wish to imply that the notion of what is "natural" has no significance for metascientific theory. One can feel it hovering impalpably around many practical research tactics such as choice of baselines, reference points, control groups and the like. But whether it has any cognitively *helpful* contribution to make to the development, application, or understanding of science, and if so, what, is still wholly obscure.[3]

In these three examples, as in the overwhelming majority of writings about scientific theory and methodology, one sees through a glass darkly when only a little polishing of the concepts employed would bring their objects into much sharper resolution. Deplorably—and incongruously—the standards of critical evaluation and pressures to clarity which have been such vital forces to progress *within* science have yet to achieve any appreciable impact on research and theory *about* science. I shall not press this complaint just now, for my immediate purpose is only to clarify what I mean by the acuity dimension of metatheory. But when later I argue for the importance of metatheoretical concomitants to substantive psychological theory, it must be understood that I presuppose a much higher level of acuity than is generally found in such work today.

I have begun with this meta-metheoretical statement about dimensions of metascience in order that I may preview the character of my remarks to follow in its terms. Part I comprises a broad survey of the nature and functioning of scientific theory, followed in part II by some normative prescriptions for practical theorizing. The substantivity of concern in these two sections lies near the extreme of abstract generality, with the result that they have a typical philosophy-of-science flavour with little content that is specifically psychological. In contrast, part III[4] illustrates the precepts of part II by close examination of selected technical components of certain recent and current behaviour theories. Acuity of analysis is a primary objective in part III, and this, unfortunately, likewise contrasts with

[3]Insomuch as it can plausibly be argued that "explaining" an event or phenomenon amounts essentially to making it appear "natural" (cf. Workman, 1964), the notion of naturalness has a legitimate place in the theory of *explanation*. But this is only one more reason for doubting that "explanation" is as such an objective of science. (If the matter were sufficiently germane here, I would argue that explanation is merely one of the pragmatic applications to which the cognitive accomplishments of science can be put.)

[4](Ed.) Not included here.

parts I and II, wherein I shall grudgingly attempt no more than a loose, largely commonsensical coverage. The reason why this discomforts and embarrasses me is that most of the traditional epistemological and semantic concepts employed by contemporary philosophy of science, including what follows below, have outlasted their technical adequacy and need to be hauled back to the shop for reworking.[5] I must slide over these inadequacies if I am to get on with anything having significance for the actual doing of psychological theory, but the result is that most of what I say in parts I and II is merely a heuristic which should not be mistaken for a technically proficient account of these matters.

I The Concept of Theory

Insomuch as our appointed purpose here is to meditate upon the role and reconciliation of theories in psychological science, it is seemly to give thought to where, within the total expanse of human endeavour, our domain of inquiry lies. That is, what is a "theory," anyway?

As is true of any word extensively deployed in ordinary language and quasi-technical discourse, the term "theory" has been and continues to be used in a wide variety of senses, some of which have only tenuous connections with the others. Since what is metatheoretically critical in one such usage need have no special relevance for another, it should be helpful to spread the entire array before us and see what issues emerge from their totality.

The distinctions which, as I sense them, inhabit the manifold uses of "theory" can best be described by a series of contrasts, some of which are "exclusive" in that one pole of the contrast is labelled "theory" to distinguish it from an opposite, while others are "inclusive" in that they express important distinctions within the term's domain of application.

Dimensions of "Theory"

A *Inclusive contrasts:*

1 Propositional theory *versus* perspectival (programmatic) theory;

2 Interpreted theories *versus* uninterpreted theories (calculi);

3 Warranted (inductively confirmed) theory *versus* speculative theory.

[5] Two examples of classical black-and-white dichotomies which many philosophers have already come to regard as a continuum of greys are the analytic/synthetic and observable/unobservable oppositions. Faced with the breakdown of traditional philosophic notions, however, far too many philosophers have retreated to the vagaries of ordinary language instead of attempting to hone their technical concepts to a keener edge.

B *Exclusive contrasts:*

1 Theoretical *versus* observational;

2 Theory (hypothesis) *versus* established fact;

3 Theory (system) *versus* isolated beliefs;

4 Theoretical (idealized) *versus* practical (realistic);

5 Theories (indicative mood) *versus* models (subjunctive mood).

A1 *Propositional theory* versus *perspectival (programmatic) theory*
In principle, the phrase "scientific theory" might be applied to any aspect—instrument, method or product—of scientific endeavour according to one's verbal whim of the moment. However, the ultimate goal of any science is to arrive at truthful (veridical) statements about the science's subject matter, and among philosophers of science, at least, it seems to be generally agreed that a scientific "theory" is a specific set of assertions about reality, or at least that the theory is a device for generating such assertions. I, too, shall adopt this position throughout most of what follows. Yet simply to stipulate that by definition, a particular psychological theory T corresponds to a specific list of declarative propositions about psychology, appears grotesquely naive when tested against what gets labelled "theory" in psychological practice. What is to be made of such familiar phrases as "behaviour theory," "gestalt theory," "statistical learning theory," "psychoanalytic theory," "S-R theory," "cognitive theory," and the like? Could one conceivably hope to characterize any one of these by a precisely specified set of propositions to which a person who accepts one of these "theories" is committed? We can easily imagine two psychologists who are both recognized behaviour theorists, or gestalt theorists, yet who are unable to agree upon a single substantive principle of psychology—in fact, a psychologist may well be a behaviour theorist or gestalt theorist without having *any* firm convictions about psychological fact. When "theory" is this broadly conceived, it no longer seems quite appropriate to think of the theory as having any specific propositional content which is true or false. (What, for example, would verify or refute statistical learning theory? The very question seems meaningless.) Instead, a "theory" in this sense is primarily a certain *perspective* on psychology, characterized by concern for a particular body of problems and phenomena, by predilection for a distinctive cluster of technical terms even though these may have no fixed usage, by prominence of certain patterns of data organization and interpretation even though what is said through their medium remains flexible, and so on for a host of features which characterize a cognitive style quite apart (or apparently so) from any particular assertive content that might be expressed within this style.

To be sure, it might be protested that use of the singular in "behaviour theory" "gestalt theory," etc., is simply a grammatical error; that the proper wording should be "behaviour theor*ies*," or "gestalt theor*ies*," in order to make clear that each of these is actually a broad category which subsumes a great many specific theories. This is a reasonable stand which helps to drive the concept of theory in the direction of propositional commitment, which is where it most profitably belongs. Even so, the fact remains that in the actual doing of psychology by psychologists, what is *called* a "theory" is more often a configuration of attitudes and special concerns than it is a set of specific beliefs, and that it is frequently very difficult to tease out what if anything a particular theorist actually considers to be true of his subject.

A2 *Interpreted theories* versus *uninterpreted theories (calculi)*

I claimed above that philosophers of science concur that scientific theories are, or at least correspond to, a set of propositions about the way the world is put together, but strictly speaking this is not quite correct. In abstract metatheory, it is not uncommon to find the term "theory" implicitly or even explicitly restricted to just the formal skeleton (schema) of a propositional system, while the flesh of semantic meaning which must be attached to the schema's dry syntactic bones if the theory is to have factual content is known as an "interpretation" of the theory. More specifically, the theory proper (i.e., the uninterpreted theory) according to this usage is a set of patterned concatenations—"postulates"—of abstract elements together with rules for deriving still other element concatenations—"theorems"—from them; while the theory is "interpreted" by adopting a particular co-ordination of the theory's elements and concatenation patterns with meaningful words and syntactical structures, respectively, in such fashion that (i) each postulate and theorem of the theory is thereby co-ordinated with a grammatically well-formed and cognitively meaningful assertion about the science's subject matter while (ii) the rules of derivation correspond to valid principles of logical deduction.[6] As an example, consider the following formal system (abstract calculus) U.

Uninterpreted Theory U

Elements: A symbol s is a *type-1* element of U if and only if s is either (a) $\infty, \bigtriangledown, \diamond$ or $\perp$; or (b) s is of form $\tilde{x}$, where x is a type-1 element of U. A symbol s is a *type-2* element of U if and only if s is of form $\dot{x}$, where x is a type-1 element of U.

[6] I have stated the notion of "uninterpreted theory" in a form more extreme than any actual usage I have encountered outside of mathematical logic, though Campbell (1920) and Carnap (1956) come close to it. More often, a writer (e.g. Nagel, 1961) will start by defining "theory" as formal calculus plus meaning, but then slip over into talking about the calculus (or the calculus with syntax and logical terms interpreted only) as though it is a theory in its own right. In particular, as soon as one begins to speak of "interpreting" the theory or its terms by introduction of "correspondence rules," "meaning postulates," "co-ordinating definitions" or the like, it is implied that the theory is something to which these make an addition.

Postulates: $\dot{\infty}\tilde{\perp}$, $\diamond\tilde{\nabla}$, $\tilde{\diamond}\perp$.

Derivation rules: Symbol concatenation c is a theorem of system U if and only if (a) c is a postulate of U; or (b) there exist type-1 elements x and y of U such that xy is a theorem of U and c is $\tilde{y}\tilde{x}$; or (c) there exist elements x and y, and a type-1 element z, of U such that xz and zy are theorems of U and c is xy or (d) there exists an element x and theorem t of U such that c is the result of replacing $\tilde{x}$ in t with x; or (e) there exist elements x and y of U such that $\dot{x}y$ is a theorem of U and c is $\dot{y}x$.

Theorems (*inter alia*): $\nabla\perp$, $\dot{\diamond}\infty$, $\dot{\infty}\tilde{\nabla}$.

While "theory" U entails a number of formally interesting consequences, there is no point in asking what they signify, for at this stage the theory means nothing at all. However, suppose that we construct "interpreted" theory U_1 by augmenting U with the following set of "co-ordinating definitions" ("correspondence rules").

Interpretation U_1 of U

∞ : "the class of persons troubled by existential anxiety"
∇ : "the class of neurotic persons"
$\diamond$: "the class of well-adjusted persons"
$\perp$: "the class of persons in need of psychotherapy"
xy : "x is included in y"
$\tilde{x}$: "the class of all persons not in x"
$\dot{x}$: "a nonempty subclass of x".

Under this interpretation of U, its postulates hypothesize that

> some persons who are troubled with existential anxiety do not need psychotherapy ($\dot{\infty}\tilde{\perp}$);
>
> no well-adjusted person is neurotic ($\diamond\tilde{\nabla}$);
>
> all maladjusted (i.e., not well-adjusted) persons need psychotherapy ($\tilde{\diamond}\perp$);

its derivation rules are valid (i.e. any interpreted theorem must be true so long as the interpreted postulates are true), and among its theorems are

> all neurotics need psychotherapy ($\nabla\perp$);
>
> some well-adjusted persons are troubled by existential anxiety ($\dot{\diamond}\infty$);
>
> neurotics aren't the only persons troubled by existential anxiety ($\dot{\infty}\tilde{\nabla}$).

However, U_1 is just one of an unlimited number of alternative interpretations for U, e.g.

Interpretation U_2 of U

∞ : "the class of all minerals that dissolve readily in water"
$\bigtriangledown$: "the unit class of mineral sodium chloride (salt)"
$\diamond$: "the class of all minerals that ionize poorly in water solution"
$\perp$: "the class of all minerals that conduct electricity well in water solution"
xy : "x is included in y"
$\tilde{x}$: "the class of all minerals not included in x"
$\dot{x}$: "a nonempty subclass of x".

Under interpretation U_2, the postulates of U translate into the assertions that

> not all minerals that dissolve readily in water conduct electricity well in water solution ($\dot{\infty}\tilde{\perp}$);
>
> salt is a good (i.e. not poor) ionizer in water solution ($\diamond\tilde{\bigtriangledown}$)
>
> all minerals that ionize well (i.e. not poorly) in water solution are good conductors of electricity in water solution ($\tilde{\diamond}\perp$);

while the theorems of U_2 corresponding to the ones cited for U_1 become

> salt conducts electricity well in water solution ($\bigtriangledown\perp$);
>
> some minerals that ionize poorly in water solution readily dissolve in it ($\dot{\diamond}\infty$);
>
> salt is not the only mineral that conducts electricity well in water solution ($\dot{\infty}\tilde{\diamond}$).

According to the formal-schema sense of the term "theory," then, belief systems U_1 and U_2 both incorporate the same theory, namely U, but interpret it differently through adoption of different sets of co-ordinating definitions for the theory's ingredients. This meaning of "theory" is a perfectly respectable one in mathematics and formal logic, for a theory's semantic content is quite irrelevant for the structural properties which interest the logician, and realization in depth that a given formal calculus can submit to a multiplicity of interpretations was one of the epochal achievements upon which modern mathematics is grounded. Introduction of this usage into scientific metatheory, however, is much harder to justify. As the term "theory" is actually used in science, it would be intolerable to claim that a certain set of beliefs about existential psychotherapy (U_1) is a manifestation of the *same theory* as certain beliefs about electrochemistry (U_2) merely because they share a common formal structure. Within a science, what is important is not what the theory may be like as an uninterpreted calculus but what it asserts; and its isomorphism, if any, to other structures elsewhere is wholly irrelevant to the theory's factual significance. Except where explicitly indicated, therefore, I shall henceforth use the term "theory" only in the sense of *interpreted* theory, namely, a cognitively meaningful assertion about reality.

In contending that the formal-calculus sense of theory has no value for science, I wish also to express doubts about the utility of views wherein a theory, though defined to include meaning, is thought to be usefully reconstructed for metatheoretical purposes as comprising (i) a syntactic calculus and (ii) a set of correspondence rules or co-ordinating definitions which give the calculus its semantic padding. With one dubious exception to be described in a moment, the calculus/co-ordinating-definitions partition is entirely otiose. The product of all this fancy formalistic footwork is simply a set of semantically meaningful propositions with cognitive implications, and with very few exceptions, this is the only form in which theory ever originates or gets used in science. I have no desire whatsoever (in fact quite the contrary) to deny that formalization is immensely helpful for explicating the logical interrelations among statements. However, to suggest that a theory's syntax is in some illuminating philosophical sense independent of, or prior to, its semantic meaning, and that in deductions made from the theory the latter is just carried along for the ride, is a dangerously backwards way of looking at the matter (see p. *207ff*). Metatheoretical partitioning of a theory between formal calculus and co-ordinating definitions clarifies its cognitive status only so far as such a partition brings insight into the cognitive status of *any* set of assertions, whether these compose a "theory" or not. But across-the-board partitioning of this sort is impossible—each co-ordinating definition would itself need to be partitioned between syntactic form and higher-order co-ordinating definitions, thus precipitating an infinite regress—and I see no virtue in submitting theories to this special indignity. While the semantic properties of theories raise some exceedingly challenging, intricate and important questions, the very same questions ultimately apply to semantically meaningful expressions of any sort; and analysis of *what* the meaning or referent of a given expression may be is obfuscated rather than expedited by the sidestep of construing the expression as a syntactic structure and a meaning glued together by correspondence rules.

The one place where a case can be made for metatheoretical reconstruction of theories as formal calculi *cum* co-ordinating definitions is in the thesis that the nature of "theoretical terms" (see B1, below) can be explained through the concept of *partial interpretation.* By a "partial" interpretation of a formal calculus is meant an assignment of semantic significance to some but not all of its elements and structures. For example, let interpretation U_1^* of U be just like U_1, above, except for omitting the correspondence rule that $\diamond$ is to be translated as "the class of well-adjusted persons." (U_1^* doesn't substitute any *alternative* translation of this symbol; it simply fails to stipulate any meaning for $\diamond$ at all.) Then U_1^* is a partial interpretation of calculus U, and the nasty problem which arises is how to characterize U_1^*'s cognitive status. It continues to generate the very same theorems as before, but now only some of these are translated into meaningful assertions, namely, only those not containing $\diamond$, even though most of these could not be deduced were the $\diamond$-axioms to be deleted from the theory. Hence U_1^* cannot, *prima facie*, possibly be *true*, for this would require that each of its postulates be true, contrary to the apparent meaninglessness of $\diamond\tilde{\nabla}$ and $\tilde{\diamond}\perp$. By the same token, neither can U_1^* be semantically *false*, for this, too, presupposes that all of its postulates are cognitively significant. (We would hesitate, for example, to say that the conjunction, "Grass is red and some blehews farble", is *false*, even though its first component is false.) For this reason, practising scientists of a meta-theoretical turn of mind have often averred that theories which cannot be completely stated in observational terms are neither true nor false but only more or less useful. Even before discussing the observation-language issue, however, it may be observed that the concept of "partial interpretation" contributes little if anything to our understanding of a theory's cognitive significance. For if adoption of a "partially interpreted" theory does, in fact, leave its uninterpreted terms meaningless, then the theory may be construed as simply a device for generating those of its theorems for which interpretations are provided, and the theory is then cognitively equivalent to (and is in this sense semantically true or false as a truth-function of) the set of its interpreted theorems. Alternatively, if adoption of such a theory also gives meaning of some sort to its "uninterpreted" terms, the doctrine of partial interpretation does nothing to illuminate what this meaning might be or what factual commitments reside in the theory's "uninterpreted" theorems. In fact, the "partial-interpretation" concept is actively misleading under this latter possibility, insomuch as if assertion of a theory containing terms not supplied with explicit observational meaning does give semantic significance to them, then there remain no semantically meaningless expressions in the theory in virtue of which its interpretation is only "partial."

Before we turn altogether away from the usage in which "theory" is a formal structure with adjustable meanings attached, it should be mentioned that, in practice, theories often have a character which makes their application to various specific cases falsely appear as though the theory were but a calculus which receives different meanings on different occasions from co-ordinating definitions which vary from application to application. This occurs when the main predictive force of the theory resides in postulates or theorems to the effect that if certain observable entities are of kinds (i.e., have properties) $\tau_1, \ldots, \tau_n$ then these entities will behave in such-and-such a way, but where the theory does not specify a set of antecedently observable conditions which are logically sufficient for something to be of kind τ_1. Then before it can be concluded on a particular occasion that according to the theory, entities $e_1, \ldots, e_n$ should behave in such-and-such a way, it must first be additionally surmised that $e_1, \ldots, e_n$ are, in fact, of kinds $\tau_1, \ldots, \tau_n$; and the auxiliary hypothesis that e_i is an *instance* of τ_i is easily mistaken for the "co-ordinating definition" that in this particular application of the theory, e_i *is* (i.e. is to be identified with) τ_i.[7]

A3 *Warranted (inductively confirmed) theory* versus *speculative theory*
This distinction, which is seldom properly appreciated either in metascience or science proper, is, of all the issues touched upon in this paper, probably the one with the greatest practical import. However, I defer its introduction until discussion of contrast B2, below, and its elaboration until the end of Part II.

Bl *Theoretical* versus *observational*
Most renowned of all the problems surrounding the epistemological credentials of scientific theories is the one addressed by the sense of "theoretical" which contrasts with "observational." What is intended here is primarily a distinction between *terms* (i.e. words or, more precisely, concepts) though derivatively, a sentence is also said to be "theoretical" in this sense if it contains one or more theoretical terms. As the observational/theoretical dichotomy has been traditionally formulated in empiricist metascience, there exist certain entities—objects, attributes, relations and the events they constitute—which we are able to experience directly, or *observe*. (Whether what is so observed are phenomenalistic sense data, ingredients of the more distal commonsense world, or perhaps something still else again, is moot but here irrelevant.) If such an entity has, in fact, been observed by us, then we can add to our vocabulary an "ostensively defined" term which designates (represents, symbolizes, refers to, is about) this entity. (*How* ostensive definition is able to endow terms with meanings that enable them to refer to observed entities is likewise moot and, fortunately, likewise here irrelevant.) Logical terms—i.e., words such as "and," "or," "not," "some" and "all"—and expressions which can be explicitly defined out of logical and previously introduced observational terms are also considered to be "observational." Then any meaningful word in our total vocabulary which is *not* observational in one of the ways just described is by definition a "theoretical" term, and the agonizing epistemological perplexity which then arises is: what do theoretical terms designate and how do they acquire their meanings? *Logical positivism* (which today has virtually disappeared from the philosophical scene) solved this tidily by denying the existence of theoretical terms altogether. In the positivist view, visual or auditory forms which behave syntactically like words but are not observational expressions simply have no cognitive meaning at all; and the sentence-like structures in which they occur (e.g. the postulates of a partially interpreted calculus) are no more than calculation devices which can be used to generate observation-language consequences but semantically assert nothing in themselves. In contrast to this extreme position, *Logical*

[7]The distinction between a theoretical kind τ_i and its empirical (observational) instance e_i is presumably what Koch (1954, p. 28f.; 1959, p. 739f) is trying to draw by his contrast between "systematic variables" and "experimental (empirical) variables."

Empiricism (which is still very much with us and to which I confess my own allegiance) insists that cognitively meaningful theoretical terms can and do exist, and that they acquire their meanings from the observational connections given them by the theory in which they are imbedded.[8] In particular, according to this view, while the truth or falsity of an observation-language statement (i.e., a statement constructed wholly out of observational terms) is generally verified or refuted by direct observation of the entities named therein, statements containing theoretical terms can be confirmed or disconfirmed only indirectly by inference from accepted observation statements and presuppose the truth of the theory by which the theoretical terms are introduced. To my knowledge, no third alternative to these two theses—positivistic and empiricistic—about how words which allegedly designate unobserved entities acquire their meanings has ever been proposed, though many philosophers have apparently felt free to reject both without offering any substitute theory of meaning.

The simplistic classical conception of the observational/theoretical distinction as a black-and-white dichotomy has generally fallen into disfavour among contemporary philosophers, partly because the difference between observability and non-observability can plausibly be regarded only as a matter of degree rather than of kind (cf. Maxwell, 1962), but more importantly on grounds that there is no such thing as pure observation. In particular, according to a recently emergent viewpoint which I shall refer to as the "omnitheoretic" thesis, *all* (non-logical) concepts used by a science at any given stage of its development are imbued with meanings given to them by the science's theoretical commitments at that time (Hanson, 1958; Feyerabend, 1963; Kuhn, 1962) and hence, contrary to empiricist doctrine, there are no propositions that the science can verify or refute by direct observation independent of the observer's theoretical beliefs. This view assuredly has *some* truth in it, for even familiar perceptual concepts in whose application we feel most secure turn out, under close analysis, to contain implications which go far beyond anything which can be conclusively verified on the specific occasions of their attribution.[9] Thus when I see that the basketball on the court before me

[8]The thesis that theoretical terms acquire cognitive significance by bumping up against observational terms in an accepted theory has occurred repeatedly in the recent philosophical literature (cf. Feigl, 1956, p. 17f.). *How* this acquisition of meaning takes place, however, has nowhere been spelled out; while what it is that theoretical terms *refer to* has, to my knowledge, been seriously addressed in only two publications, an unnecessarily abstruse monograph by myself (Rozeboom, 1962) and a short article by Carnap (1961); criticized in Rozeboom (1964). It is sometimes claimed that a theory gives meaning to its nonobservational terms by "implicit" definition (e.g. Nagel, 1961), but the traditional theory of implicit definition proffers only necessary, not sufficient, conditions for the designata of terms so defined—whatever an implicitly defined term refers to must be something which satisfies the definition—and hence fails to say what an implicitly defined term *does* designate.

[9]Contrary to the impression given by recent advocates of this view, empiricists of an earlier era were well aware of the "surplus meaning" in ordinary-language concepts (e.g., Ayer, 1936,

is a *ball*, my perception includes, among other things, a judgment that this object is also convex on the side turned away from me. A more interesting example with practical relevance for psychology is the concept of "behaviour." Surely we should be able to tell by observation whether or not a given organism is behaving, or else what could justify the behaviouristic thesis that behaviour and environmental events are the data of objective psychology? Yet suppose that we are watching a play and see one of the actors suddenly fall to the stage. Is this behaviour or not? Ordinarily we would say that it is, especially if this event fits the play's dramatic integrity. But what if we had seen a falling ballast weight strike the actor a crushing blow on the head just before he collapsed: Would we still think of his floorward movement as *behaviour*—i.e. is this something he *did*? Certainly not—but why should information about the blow affect our judgment? A little reflection on this and similar examples reveals that "behaviour" is not *merely* a change in the spatial loci of a living organism's bodily parts, but also requires that the change have a certain special kind of underlying cause.[10] Thus to identify an organism's motions as *behaviour* is to make a judgment far richer in theoretical commitments than anything which could reasonably be called a *pure* observational report.

Realistically, then, it must be admitted that theory-free observation is a limiting ideal which scientific practice never completely attains. Even so, as soon as the omnitheoretic thesis plunges beyond this to allege that no significant difference exists between observation and theory, it becomes a retrograde mystique which has lost intellectual contact with the actual doing of science. As anyone who has ever engaged in serious research is acutely aware, the distinction between facts of which we are most certain—i.e. *data*—and the inferences which we attempt to draw from them is of the utmost methodological importance, and a goodly proportion of the technical aspects of "scientific method" have been developed precisely to keep the perceptual beliefs with which scientific inferences begin as sharply distinguished from the conclusions to which they lead as is humanly possible. Even if none of our observational concepts, scientific or otherwise, are altogether free of theoretical overtones, it does not follow that they prejudge the veridicality of *every* theory to which they are evidence-wise relevant, and it is simply not true, even in ordinary life much less in technical science, that we can never perceive and describe an event without committing ourselves to one or another of rival theoretical interpretations of it. In the example of the collapsing actor, for example, we can perfectly well see that he has fallen to the floor without any commitment (if we make a modicum

Ch. 7) and for this reason argued that what we *really* observe directly are sense data.

[10] Just what sort of underlying cause is envisioned, however, is intriguingly muddy. Commonsensically, calling a bodily movement "behaviour" implies that it was done on purpose. But this would exclude such actions as heart-beating, pupillary responses and involuntary postural adjustments which are unquestionably—and importantly—included in the technical behavioristic concept of "behaviour."

of effort to abstain from hasty judgment) as to whether or not he was engaging in falling *behaviour*. In actual behaviouristic research, the experimenter usually makes a special point of recording and reporting his data in such fashion that he and his readers can agree about what happened external to his subjects' skins *without* necessarily agreeing about what went on inside. In particular, whether or not the physical motions or environment alterations effected by an organism are instances of *behaviour* makes not the slightest difference for the researcher's protocol record—irrespective of whether, e.g., a rat pushes the lever with grim determination exuding from his clenched jaws and hunched shoulders or just bumps it accidentally with his tail, the response recorder tabulates exactly one pulse from the lever relay. It is precisely *because* behavioural research has become able (albeit still imperfectly) to think about an organism's overt motions and their environmental consequences without *presupposing* any particular underlying cause that behaviour theory has been able to push beyond commonsense mentalism toward a genuine science of psychology. And this is true throughout all of scientific research. Whereas proponents of the omnitheoretic thesis have emphasized the theoretical overtones of slovenly, uncritical concepts drawn from everyday life or the physical theories which underlie interpretation of instrument readings, the former are the very sort of notions that science is most insistant upon replacing with purified technical counterparts, whereas in the latter case most research scientists are not at all reluctant to acknowledge that their theory-dependent instrument interpretations *are* uncertain inferences, and are perfectly able to question the soundness of these inferences without disturbing their perceptual beliefs about what the instrument readings themselves are. I am, admittedly, arguing this point more by loud shouting than by careful analysis of real-life examples, but I submit that anyone who proposes that empirical scientists do not really have an observation language, in which are couched perceptual beliefs that are independent of the theories upon which these data are brought to bear, simply does not appreciate what goes on in actual research. By no means do I wish to imply that this expurgation of theory from data is easily accomplished, or that scientists always maintain high standards in this respect.[11] This is an unceasing challenge to both the science's technical development and the individual investigator's professional skill. But within any empirical science there is an unrelenting *press* toward exhuming and casting out from the science's technical data language any presupposition which, upon critical contemplation, will sustain reasonable doubt—and another way in which "scientific method" leaves everyday habits of thought far behind is in development of professional proficiency at reasonable doubting.

[11] I have myself called attention to ways in which the orthodox wording of data summaries in behavioral research has prejudged importantly problematic theoretical issues (Rozeboom, 1958, 1960). But rather than contravening my present point, the fact that I *was* able to show this demonstrates that it is not *necessary* to read these inferences into behavioural observations—we can agree about the data even while disputing their interpretation.

The fact that perceptual judgments, even in scientific data gathering, never attain perfect philosophical incorrigibility, but buy into a broader framework of organized beliefs, complicates analysis of the observational/theoretical distinction but in no way demeans its scientific significance. Whatever is to be made of it philosophically, scientists have *de facto* observation languages within which they formulate what they *take* to be their science's data. However corrigible and theory-saturated (on one level of theory) these professed datum-judgments may be, they are still the bases on which all further inferences in the science are grounded, and the words used to express them are, by definition, the science's observational terms. It is a further undeniable linguistic fact that in most sciences, many of the words which occur in sentences that are seriously (if tentatively) entertained by the science in explanation of certain data or judged for plausibility on the basis of their relation to datum-beliefs are neither observational terms nor are analytically reducible to observational concepts. Words of this sort are then by definition *theoretical* terms in the sense that contrasts with "observational," while any proposition containing a theoretical term is likewise "theoretical" in this sense. Thus the observational/theoretical distinction remains an obdurate fact of scientific inference with which metascience must reckon, irrespective of how it is to be interpreted epistemologically, and it would be a disaster for scientific practice and metascientific acuity were persuasive proponents of the omnitheoretic thesis allowed to undermine our appreciation of its methodological importance.[12]

B2 *Theory (hypothesis)* versus *established fact*
Another way in which the terms "theory" and "theoretical" are commonly used is to stigmatize the propositions to which they are applied as uncertain or hypothetical, in contrast to propositions whose truth-credentials are so strong as to admit of little practical doubt. Thus we say, "That's pretty theoretical," to express our feeling that a certain surmise or speculation has only a tenuous grounding on established facts—not that the facts necessarily *impugn* the proposition in question, but only that they do not strongly support it. Subjunctive considerations incorporating an improbable antecedent are also frequently characterized as "theoretical" in this sense as when, for example, we theorize about what would happen on Earth were the sun's energy output to decrease one per cent, or speculate about how

[12]Most readers who are themselves professional scientists will be puzzled why I have given so much space to arguing this point, which is standard doctrine in the theory and practice of research. Unfortunately, the legitimate dissatisfaction which has burgeoned among philosophers and historians of science over traditional reconstructions of scientific methods and concepts is beginning to flow most vigorously into the omnitheoretic stream. To date, the latter is still but a newly started freshet, but it could easily become a torrent precisely *because* it is a bold rejection of past orthodoxies whose occasional shoals and silt accumulations are becoming increasingly intolerable to modern high-powered, deep-draft philosophical navigation. (To belabour the metaphor further, I urge that we put some of these high-powered modern resources to dredging instead of abandoning them and the main empiricist channel for a wilderness portage.)

we could most profitably invest our winnings if our ticket in the New Hampshire sweepstakes just happened to hit a big pay-off.

While there is considerable overlap between applications of the term "theoretical" in its non-observational (B1) and hypothetical (B2) senses, the congruence is by no means perfect and it is important that the distinction between these two meanings remain fully appreciated. In the first place, theories in sense B2 need contain no theoretical terms in sense B1. For example, if planetary positions recorded on particular occasions are construed to be astronomical "data," so that astronomy's observation language contains spatio-temporal co-ordinates and the names of planets, the celebrated hypotheses of Copernicus and Kepler are observation-language theories involving no theoretical terms. Again, serious surmise that certain puzzling geological formations were produced by glacial erosion, or that the human species will lose all of its bodily hair within the next million years, is a B2-type theory essentially at the observational level. (I qualify "surmise" with "serious" here because there is a tendency to withhold the title "theory" in sense B2 from hypotheses which are no more than arbitrary guesses.) These theories are "observational" not because they have been or can be conclusively verified by present observation, but because their conceptualization requires us to appeal to no aspects of reality other than what is already perceptually known to us.[13] Similarly, any plausible but inconclusively established generalization about the observable properties of observable things (e.g., extrapolation from an observed sample distribution to a population considerably more inclusive than the one sampled) is a "theory" in sense B2. On the other hand, a term or proposition which is "theoretical" in sense B1 need not be so in sense B2. It is possible for assertions containing theoretical terms to have such strong evidential support that they no longer sustain appreciable doubt. For example, a great deal of chemical theory concerning the atomic constituents of macroscopic substances must surely be regarded by chemists who well understand it to be about as certain as any generalization, observational or theoretical, ever gets, and the same could be said for much of the gene theory of heredity. Similarly, while reflexes and associations are unobserved (i.e., theoretical$_{\mathrm{B1}}$) entities postulated by psychologists to explain certain stimulus/response correlations observed at various times in the lives of certain organisms, the bare existence of reflexes and associations can scarcely be questioned even if many of the more elaborate behavioural theories which have been built upon associationistic concepts well merit continued scepticism. More generally, science and everyday life abound with reference to *dispositional* attributes which

[13] Note that while the hypothesis, e.g., "Marsupials were common in America 50 million years ago," envisions a situation which is no longer itself observable, this proposition is constructed out of concepts that refer only to attributes which we have commonsensically "observed." Thus if we can consider ourselves to have perceptual awareness of brief time intervals, say one-second durations, and of the temporal-precedence relation, we can define the complex attribute *being 50 million years prior to the present* wholly in observational terms.

are "theoretical" in sense B1 (cf. Rozeboom, 1961, p. 362ff), yet whose reality—though by no means everything else surmised about them—is so undeniable that there is considerable reluctance to regard them as theoretical at all.

In all these examples of well-established theoretical entities, I have deliberately called backhanded attention to participation of the terms which refer to them in more speculative hypotheses as well as in near-certain beliefs. As discussed more thoroughly in part II, scientific theories are not monolithic wholes which stand or fall in their entirety, but are susceptible to analysis into conjunctive components each of which has its own degree of credibility. Disentangling the components of a theoretical complex which are *warranted* by the evidence at hand from those which are largely unsupported speculation has an importance for practical theory building which cannot be overestimated, and for this reason, if no other, it is essential to be clear that theoretical$_{\mathrm{B2}}$ (*uncertain*) does *not* entail theoretical$_{\mathrm{B1}}$ (*nonobservational*).

B3 *Theory (system)* versus *isolated beliefs*
Whereas any aspect of a theory, great or small, may be characterized by the adjective "theoretical," the noun "theory" carries the implication that whatever is so categorized is the whole thing. (Cf.: whereas "my" applies to anything that is mine, "I" refers to all of me as a unit.) Thus while individual words and singular assertions may be "theoretical," we are not likely to call them theories in and of themselves. If we ask what sort of whole is to be dignified by the label "theory," moreover, we feel that whatever else it may be—propositional or perspectival, observational or nonobservational, warranted or speculative—it should have breadth and potency, that it should provide a conceptual framework within which not just one or a few but a great mass of specific facts and possibilities are organized into an intellectually workable pattern. This is why a simple restricted generalization, even if sufficiently uncertain to count as "theoretical" in sense B2 (e.g. hypothesizing on the basis of a sample of size 10 that the linear correlation in a particular strain of dogs between a certain measure of tactile sensitivity and a certain measure of emotionality is about .50) would not generally be considered a *theory* in any meritorious sense, whereas a system of interlocking generalizations of which this is one (e.g. many observed correlations in many different species between various measures of sensory function and motivational arousal) might be subsumable under a few higher-level generalizations which would be admitted to the class of theory with much less hesitation. (It is almost impossible to keep theoretical$_{\mathrm{B1}}$ concepts from intruding into such higher-level generalizations, but it is not logically essential that they do so in order for the system to qualify as "theory"—e.g., the Copernican and Keplerian theories of planetary motion cited earlier.) This use of the term "theory" to denote a pattern of organizing ideas is nowhere more apparent than in its application to mathematical systems such as

Information Theory, Multiple Regression Theory, and the like. These are neither uninterpreted postulate sets (cf. A2), assertions about unobserved entities (cf. B1), nor indeed hypotheses (cf. B2) of any sort. Instead, they are systems of explicit definitions, and the analytic truths which follow from them, which enable their user to perceive within an otherwise bewildering complexity of raw data an abstract comprehensible empirical structure.

It should be added that while the term "theory" occurs rarely if ever in scientific parlance without connotations of "system" or "conceptual organization," this is by no means true more generally. Historical explanations, in particular, are frequently theories in sense B2 but not in sense B3. Thus when the police chief announces that he has a theory about the mysterious disappearance of Lady Wimbleton's jewellery, or anthropologists entertain theories about the genetic and cultural interactions between Neanderthal and Cro-Magnon man, what is at issue is not organizing principles but specific problematic historical events.

B4 *Theory (idealized)* versus *practical (realistic)*
Still another sense of "theory" is the one which contrasts with "practice," as when we compare what "in theory" should happen in some situation with the actual realities of the case. "Theory" as conceived here is not so much a hypothesis whose implications for the situation in question are refuted by the data as it is a *principle*, *rule* or *ideal* to which reality conforms only approximately. In everyday affairs this sense of "theory" subsumes normative as well as natural principles, as in, e.g. "In theory, drivers should come to a complete halt at all stop signs, but in practice the 'rolling stop' is common." In science, theory-as-ideal is what *ceteris paribus* clauses are invoked to protect—i.e., we say that such-and-so is the way things would be were it not for certain secondary disturbances which have not as yet been brought into our systematic account of the matter. Thus we might assert, "In theory (ideally, in principle, as a rule) an adult human has 32 teeth, though decay, accidents and developmental anomalies often cause the actual number to differ from this"; or "In theory, intralist interference in rote verbal learning increases as a function of intralist similarity, but in practice the data are much more complicated than this." In these two examples, the "theory" is not a disconfirmed speculation but an accepted trend which would be expected to hold exactly (or essentially so) were certain other variables, some known, some not, controlled in appropriate ways.

To be sure, the theory in a theory/practice distinction does not have to be inaccurate *merely* in being a simplification—in ordinary language, this contrast not infrequently carries the suggestion that the theory in question is wildly inappropriate to the hard-headed realities of the matter. But it should be emphasized that "theory" in the sense of idealization is perfectly respectable cognitively, and indeed, if buttressed by the proper qualifications, can stand without apology as an aspirant to literal semantic truth. For subjunctive and counterfactual propositions of form "If p were the case, then q would also be so" are not at all impugned by mere failure of p to be the case; and while what does verify or refute subjunctive and counterfactual conditionals is still hotly disputed by philosophers, a theory which tells us how certain variables *would* be precisely related were certain secondary disturbances properly controlled has made a commitment to the way the world is put together which may not merely be true but can also have considerable value for practical explanation, prediction and control even when these secondary effects are not eliminated. Similarly, if the imperfect dependence of a variable Y upon variables $X_1, \ldots, X_n$ under circumstances C is described by, say, the curvilinear regression of Y upon the X_i, the statement "Y is related to $X_1, \ldots, X_n$ under circumstances C by the function $\hat{Y} = \phi(X_1, \ldots, X_n)$" can be *exactly* correct *qua* regression (at least within the limits imposed by sampling uncertainty) irrespective of how much residual scatter Y may have around this trend.

B5 *Theory (indicative mood)* versus *model (subjunctive mood)*
One of the most popular catchwords in the behavioural sciences today is the term "model," not least among whose virtues is its union of technical clang with such free-wheeling ambiguities of meaning that an amateur metatheoretician can achieve instant profundity by talking about the role of "models" in science with very little danger that what he has said is wrong in every sense of the word.[14] The variegated meanings of "model" substantially overlap those of "theory"; in fact, theory-as-uninterpreted-calculus (A2), theory-as-simplified-ideal (B4) and to some extent theory-as-system (B3) have all been identified as "models" at one time or another. There is, however, one important usage in which "model" is set in opposition to "theory." Specifically, let T be a set of propositions which entail certain known or suspected empirical phenomena, or at least *ceteris paribus* idealizations of them. Then T is a model of these phenomena if we say that it is *as though* T were the case, whereas T is a theory if we hypothesize that T *is*, in fact, so. That is, when "model" and "theory" are contrasted, the difference lies not in propositional content but in the mood of speech, the model being a subjunctive expression of what the theory asserts indicatively.[15] For example, the stimulus-sampling model of probability learning hypothesizes that a molar stimulus S controlled by the experimenter behaves like an ensemble of micro-stimuli, only a randomly fluctuating subset of which confronts the subject on particular

[14] A useful summary of the major meanings of "model" may be found in Kaplan (1964, Ch. 7). See also Chapanis (1961) for an excellent discussion of "models" in what I, for one, consider to be the core sense of the term—namely, where "x is a model of y" is to be translated as "x is a system hopefully isomorphic to y in important respects," or more briefly, as Chapanis would put it, "x is analogous to y."

[15] Cf. also Boring (1957) and Workman (1964) on the subjunctive/indicative distinction between models and theories.

presentations of S; but *qua* model, the hypothesis reserves judgment as to whether or not S *in fact* has such a composition.

To attend merely to a model's iffy tone, however, is to overlook the crucial point that it, too, makes indicative commitments about the nature of the phenomena to which it is applied. For we can readily imagine disputing whether or not the phenomena really are as though T were the case. The phenomena must *be* a certain way in order for them to favour one *as-if* assertion over another. That is, to say that something is *as though* T were the case is to allege that some of T's implications *are* true. As a result, proposing hypothesis T as a "model" rather than as a straightforward theory usually has the force of saying, "I think that T is at least partially true, but I would be very surprised if it were correct in all respects and I'd rather not try to say just now which parts are true and which are not." This is a perfectly legitimate attitude to take toward a problematic hypothesis, but it raises some rather basic metatheoretical questions about what a science profits from calling a hypothesis a "model," rather than either (a) asserting it indicatively, if tentatively, as a theory, or (b) claiming only the known (or suspected) facts about the phenomena in question in virtue of which the model is reasonable *qua* model in the first place. One major benefit from looking at consequences of the model beyond merely those already known to be approximately correct is that this may well promote discovery of previously unsuspected regularities in the data, or suggest rational norms (idealizations) whose empirically determined parameters and/or divergence from the facts in particular instances turn out to define significant data variables; however, this can be accomplished just as well by considering the model indicatively as theory. In part II it will be seen that thinking of hypotheses as "models" in the present sense can be a helpful first step toward what will be argued is the proper way to do theory. But meanwhile, it should also be pointed out that overly facile insistence that a given hypothesis is *merely* a model is a form of methodological irresponsibility. For this attitude encourages one to shrug aside all protests that certain features of the model are theoretically implausible, contrary to known fact, or even logically inconsistent, with the glib rejoinder that a model doesn't claim to be literally true. It is, however, one thing to propose a hypothesis in full expectation that it will not prove tenable in all respects, and quite another to propose it even when it contains specific identifiable components which are strongly believed to be incorrect. A theorist (or modellist) who wishes to pursue the latter course has the burden of responsibility for showing that the use to which he is putting this model is successfully served *even though* it contains these contrary-to-fact ingredients, and to explain why revising the model to eliminate them would not serve this purpose even better.

II Metatheory: The Task Ahead

So much for the various senses of "theory." The review has relevance for a conference on psychological theory if only to promote agreement on what we are talking about. But mere facilitation of communication (of which, in my gloomier moments, I sometimes despair altogether) is not my primary purpose in citing them. Each of these themes in the meaning of "theory" has as counterpoint one or more aspects of the technical functioning of psychological theories which are critical for metatheoretical judgments about what theories should be like and how these ideals can most effectively be realized. It is to such judgments that I now turn.

Norm-free analysis of theories

Let me say at the outset that while the prescriptions I shall offer would, if effected, powerfully influence the rapprochement and development of psychological theories, their impact would come to bear in a curiously indirect way. For I want to insist first and foremost that *any psychologist should feel perfectly free, unconstrained by metatheoretical norms, to say anything that he very well wants to say, in observation language or not, precise or vague, guarded or imprudent, earnest or jocular, quantified or qualitative, hesitant or dogmatic, so long as he feels that it contributes something, however obscure, to our understanding of the phenomena under concern.* As I read him, this is essentially the message that Koch (1959) would deliver to us from his official eyrie above the expanse of contemporary psychological thought. Much as I deplore the anti-methodological undertow in Koch's more recent writings, I thoroughly agree with him that it would be foolish for psychological theorists to constrict their thinking only to those conceptual procedures seemingly vouchsafed by positivistic metascience of the '30s and '40s. Unlike Koch, I am sceptical that the doctrines imported into psychology from philosophy of science and other exogenous disciplines during the last few decades have done any appreciable harm—while there was considerable agitation of psychology's verbal surface during this period, I suggest that the only disturbances which penetrated to the level of substantive research were home-grown issues, such as the intervening variable hassle and before that the nature of psychological data, which may have paralleled and eventually made contact with philosophical developments elsewhere, but which arose indigenously out of deeply felt needs to clarify our objectives and methods. These needs (with some reorienting of their cathexes) are as strong as ever, but they will not be met by a childlike appeal to past or present philosophy for advice on how to do psychology. Virtually all philosophers of whose work I am aware stand in awe of the juggernaut of technical science and wouldn't dream of tampering with the machinery. That branch of philosophy currently known as "philosophy of science" is not a school of metascientific engineering for instructing scientists in their profession, but for the most part an approach to epistemological theory which focuses upon the cognitive methods and achievements of science in hope of learning *from* science how human knowledge advances. Thus I repeat: In the proper practice of psychological research and theory, anything goes. The extant body of knowledgeable metascientific opinion comprises mainly tentative generalizations about how science has apparently worked in the past, and apart from the theory of experimental design and some aspects of statistical sampling theory, is still unprepared to say how science *should* be run. A scientist is ever so much more likely to grope, reason, intuit, plunge or blunder his way to novel discoveries and revolutionary insights if he does what comes naturally to his professional instincts than if he is constantly looking back over his shoulder to make sure that he hasn't violated somebody else's rules for the game.

But—this bacchanalia of metatheoretical permissiveness has a cold, grey morning after, and this arrives as soon as the scientist's creative inspiration (which I shall assume is an exemplar of "theory" in some sense of the word) passes over into some form of public document. Once the raw theory is palpably before us, we need to pound, grind and wring it metatheoretically in order to squeeze out just what it actually accomplishes; to make clear the respects in which it is or is not merely a restatement of notions already familiar in previous work on this problem; to recognize what is tenable in it and what is gratuitous; to extract from it what, tenable or not, is conceptually solid and what is too vague or ambiguous to do more than emote—to make, in short, a technically detailed appraisal of what the theory is and does, in as many respects relevant to the cognitive goals of psychology as metatheory has learned to perceive. The operation I am envisioning is something like diamond mining: The first stage, without which all subsequent sophistications come to nothing, is to isolate a mass of ore within which, no matter how messy the matrix may be, there is reason to hope that diamonds are embedded, while this ore is collected in wholesale, indiscriminate, scoops that would rather gulp a ton of waste than lose a gram of gem. But rich or lean, the crude ore still requires a great deal of further processing. Not until the hard stones have been searched out, cleansed of their encumbering dross, cut to exacting shapes and polished to a high brilliance has anything of major value been achieved. Unanalysed psychological theory similarly consists of a sticky gumbo through which uncut gems are occasionally strewn, and it is the task of metatheory to refine this ore.

This analogy breaks down, however, when we come to the methods of refinement. Metatheoretical analysis is most properly conducted in the indicative, not the imperative, mood of speech; and an extant theory can be dissected right down to its last *non sequitur* and grammatical inconsistency without altering the theory in any way or even recommending that it be revised. One can, for example, point out that a theoretician's verbal definition of a certain technical term is ambiguous or at odds with the use he makes of it, or that an apparently innocuous proposition in the theory actually rests upon concealed presuppositions of controversial nature, or that a certain premise is much stronger than apparently necessary for its role in the theory, without insisting that there is anything *wrong* with ambiguities, concealed presuppositions and unnecessarily strong premises. It could well be that the explanatory power or audience appeal of the theory lies primarily in such subtleties, and the job of metatheoretical analysis is simply to *show* this, not to damn it. Even if it be agreed that certain features of the theory do fall short of optimality, it is still for the theoretician himself (or others who have assumed responsibility for making sense out of the data to which the theory relates) to judge whether these blemishes are serious enough to require attention and, if so, what should be done about them. Metatheoretical analysis of the sort I have in mind should, in fact, lead to considerable enhancement of a theory's keenness and thrust, but only because much of what the analysis discloses will be recognized by the theory-builder himself, out of his own personal sense of cognitive rightness, without coercion from extraneous standards, as features of the theory which he would like to purify or eliminate. Thus it would take a pretty callous theorist to remain entirely unmoved by exposure of a logical contradiction in his basic tenets even if contradictions are not always so disastrous to a conceptual edifice as formal logic would suggest.[16] Similarly, if it were shown that a given theory can be restructured in such fashion that certain suppositions which the theorist himself finds uncomfortable no longer add anything to what the rest of the theory implies for the theory's intended domain of application, it is rather unlikely that the theorist would not wish to make some deletions.

To be sure, normative standards are bound to spin off from metatheoretical investigations in practice. A judgment of form, "Other things equal, the more of quality Q a theory has, the closer it approaches goal G," carries with it the implication that if G *is* a goal the theory is intended to achieve, then development of the theory *should* attempt to maximize quality Q, at least to the extent that this does not impede pursuit of other desired goals; and since metatheorizing will best continue to be done mostly by psychologists themselves, cheek by jowl with their substantive concerns, there will be no shortage of strongly held value judgments to accompany metatheoretical assessments. But I want to stress once again that the metapsychologist's job is first and foremost to work out in technical detail but descriptive neutrality just what the significant properties of a given psychological theory *are* and how they relate to the various possible functions the theory might be expected to perform, and only derivatively, under another hat, to award praise and blame. The distinction I am trying to make is nicely illustrated by the contrast between the two outstanding works of metatheory which have appeared in

[16]Formally, a logical contradiction entails *every* conceivable proposition. But in practice, what has in fact been deduced from postulates containing an inconsistency may or may not depend upon this feature, so that it may be possible to eliminate the inconsistency while leaving the theory's past positive accomplishments essentially unaltered.

the psychological literature to date, namely, the brilliantly *con*structive analysis of Tolmanian expectancy theory by MacCorquodale and Meehl (1954) and the brilliantly *de*structive analysis of Hullian theory by Koch (1954). Sympathetically, sensitively and with considerable success, MacCorquodale and Meehl attempted to lay bare the logical structure implicit in Tolman's intuitively appealing but technically obscure behavioural theses. The system which they extracted was no longer pure Tolman, for it is in the nature of explication to become definite or at least to make branching alternatives visible at points which are ambiguous in the original.[17] But the MacCorquodale and Meehl reconstruction allows us to see, ever so much more clearly than before, what is *in* Tolmanian expectancy theory, leaving it for the reader to decide whether or not he *likes* this sort of theory and where it should go from here. In contrast, Koch's essay is an exhaustively detailed listing of ways in which Hull's formulations fell short of certain simplistic ideals which psychological metatheory of the '40s (including Hull's own) borrowed as norms from the prescriptive models of scientific endeavour popular at the time, accompanied by a sustained, thinly muffled shriek of condemnation. Little attention is given to *why* a theory should conform to these ideals, or what there might be of methodological interest and substantive importance in Hullian theory despite its manifest imperfections, or how seriously and in what way these blemishes interfere with what the theory could otherwise accomplish, or even whether it would require more than trivial modifications to expunge them. Koch's analysis reveals something of what Hull's theory was *not*—in particular, that it was far from the quintessence of logical rigor which many psychologists had uncritically supposed it to be and that Hull's own metatheoretical characterizations of it were inaccurate—and a number of its technical inconsistencies, but very little on the positive side about what, as a conceptual structure, it *was*, what it *did*, and *how*. (Koch frequently alludes to Hull's status in these latter respects, but either left the actual analysis programmatic in his own thinking or did not consider it of sufficient importance to share with his readers.) In fairness to Koch it must be added that his essay concluded with reservations about the value of these norms in whose light the theory was found wanting, and since then he has come even more sternly to abjure this style of metatheorizing. But it would be most unfor-

[17]It is worth noting here that clarification of an ambiguity does not require commitment to one detailed alternative to the exclusion of others. The explication may amount precisely to making clear what the theory leaves open. For example, a learning theory which speaks of the growth of habit as a function of learning trials without saying *how* habit grows with trials would be clarified by stating that the theory construes this function to be an adjustable parameter, and illuminated even more by spelling out what dimensions of variation the theory envisions for this parameter—e.g. does or does not the theory allow it to assume different values for *inter alia* different habits, different learners, different periods in the life of the same learner? There is no demand that the theory supply numerical details for this parameter or even that it impose restrictions on its manner of variation. One merely wants to know where, and in what way, the theory is *intentionally indefinite*.

tunate if repudiation of *bad* metatheory were overgeneralized into indiscriminate disregard for the deepened understandings and sharpened concepts which emerge when metatheorizing is done incisively but sympathetically.

My emphasis that psychological metatheory must concentrate first and foremost upon norm-free analysis of what psychological theories are in fact like is in no way symptomatic of any reticence or lack of conviction on my part about what is ultimately desirable in a theory. The overriding difficulty is that *metascience simply doesn't know enough as yet to make useful recommendations for the technical elaboration of substantive psychological theory.* "Technical" is the operative word here. If there is anything at all to be learned from the history of science and technology, it is surely that development and application of science in all its phases turn most pivotally upon painstaking exploitation of ever more exacting detail, both in the controlled sensitivity of observations and in the refinement and elaboration of concepts. The greater the demonstrable power achieved by a scientific discipline, the greater the disparity between the articulation and precision of its cognitive machinery and that of everyday life.[18] Inevitably, then, if metatheory is to get at the wellsprings of potency in scientific thought, still muddy though these may be, it will be necessary to study the fine-grained structure of its conceptual mechanisms with a technical precision and subtlety at least as great as that of the science itself. But extant metascience has scarcely begun to achieve any significant technical acuity.[19] About the only developments approaching the degree of precision requisite for metatheoretical assertions to have determinable truth-conditions have occurred in a *few* sectors (e.g., the work of Hempel and Carnap) in which the object-theories studied are highly idealized with no more formal structure than can be comfortably formalized in the propositional calculus with an occasional touch of elementary qualificational logic. In the psychological literature, metatheory has consisted almost entirely of grandiose generalities which hover as word mists far above the concrete realities of theory-building practice, free from any recognizable meaning ties to the metapsychological *data* which should be, but virtually never are, adduced to support them. I have already made reference in my

[18]I know of no better way to appreciate this than to compare a few issues of *American Scientist* and *Scientific American*. Both are written for an all-purpose audience, but the former presupposes a modicum of scientific literacy while the latter does not, and the difference in conceptual feel is striking. (Both contain much that I, for one, do not understand, but the way in which my understanding falters is altogether different in the two cases. One shows me the mountain which I am unprepared to climb; the other leaves me groping for direction in a cloud of feathers.)

[19]Throughout this paragraph, I am using "metascience" and "metatheory" in the sense in which "-science" and "-theory" connote some degree of generality and systematization. That is, I do not intend my present remarks to extend to the frequently proficient critiques that appear in the scientific literature on restricted points of specific substantive theories. I am concerned here primarily with metatheory whose level of abstraction is high enough that what is said has relevance (or would have relevance, were it to have sufficient acuity) for more than just a single aspect of one particular theory.

opening remarks to the inadequate acuity of most metatheory, and I hope you will grant at least provisionally the accuracy of this appraisal without demanding that I produce an overwhelming documentation of it at this time.[20] There is nothing particularly surprising or shameful about this condition. It is simply a part of the human predicament, a legacy of brute inarticulateness which mankind is slowly but successfully casting off, and I cite it at this point only to emphasize that the most pressing task for psychological metatheory today is not to teach but to learn—above all, to learn how to discern, to formulate, and to think clearly about the technical refinements which give modern psychological theories whatever cognitive superiority they may possess over commonsense psychological intuitions. To do this, we shall have to trace out in exhaustive detail the conceptual structure found in psychological theorizing as it actually occurs in the systematization of data and evolution of hypotheses at the research level (the difference between working theory in serious research and what is found in secondary sources being something like Mozart *versus* Muzak); and the fact that most of this structure has never been made explicit in the professional literature, but must be dug out of vast, untidy heaps of ellipses, grammatical absurdities, sentence-to-sentence shifts in meaning, and all the other frailties of language in use, only makes the task that much more demanding of determination and skill. This is not a job which can profitably be left for professional philosophers of science. Only a person who has lived the technical concepts and problems in a particular substantive area deeply enough to have acquired a gut-feel for what theories of this material are trying to do has much hope of separating what is vital in them from what is not, of empathizing the undercurrents and overtones that turn non sequiturs into enthymemes, and of working out formalized reconstructions which are faithful more to the spirit than to the letter of the originals.

Moreover, the resources of professional philosophy are still not, as they now stand, sufficient for this job. The philosophy of cognition has sought through the centuries to abstract from human reasoning those patterns of thought which still retain normative appeal when stripped to their bare bones of logical form; but whereas the past 100 years has supported a remarkable surge of perspicacity into deductive logic, philosophical command of other belief-governing formal linkages among complex ideas is as yet little more developed than was deductive logic in

[20]It fills me with anguish that, try as I may, I can find no words which can help others to see, with the vivid awareness of first-hand experience, the putty-and-wet-sand quality of most of our conceptual tools, not merely in metatheory but in large sectors of substantive psychology as well. (Polemics such as these make evident that I hold this conviction, but do little to demonstrate its truth.) The only way in which this communication can be effected is through an amassing of concrete examples in which flabby concepts in common use are toughened up. This is an enterprise at which I have persisted in most of my previous writings and will continue in part III below. (Ed. Not included here for reasons of space.) But what can be accomplished by any one lunge is minute in comparison to the immensity of the task.

Aristotle's day.[21] It is all the more important, therefore, that metapsychology now concentrate upon exacting *descriptive* analysis of technical detail in working substantive theories, in full expectation that in order to identify and express what is there we may well have to pioneer new skills and concepts in logic, epistemology and semantics. But this is also a way—and an important one—of developing psychology proper. For as these new or sharpened methodological tools are put to practical use, the architects, builders and renovators of substantive theory (from whose peerage no psychologist should be excluded simply because he practices serious metatheory as well) will be thereby availed of keener awareness of what they are about, in what ways their theoretical products fail to be and to do what their creators intend of them, and how their theory-construction techniques can be honed into a more incisive instrument for these intentions. In short, I conceive of future metapsychology as a technological service adjunct to psychological research, staffed largely by professional psychologists with specially trained skills at conceptual analysis, whose utility for substantive psychological theory is comparable to that of inferential statistics and the theory of experimental design for data collection and of electronics technology for the refinement of research hardware.

Metatheoretic issues: Some examples

To be sure, I have sketched this vision of metapsychology's lusty, trail-blazing future in wispy impalpables such as dreams are made of, and until you are given some operational specifics you have every right to remain sceptical. Let me, therefore, review some of the more urgent metatheoretical problems which protrude from the multiple meanings of "theory" surveyed earlier, and conclude (part III) with a sampler of metatheory in action.[22]

To continue, then, recall that a great deal of what passes for psychological "theory" actually appears to contain little definite propositional content (A1). But this is a remarkable thing. How is it possible—or is it?—for a point of view to have cognitive force and yet not assert anything? In particular, if competing perspectives do not incorporate incompatible beliefs about psychological fact, in what sense can there be conflict among them? To the extent that perspectival theories differ merely in what they are about, as on a grander scale the subject matter of physics differs from that of sociology, or in the murmur and clang of their preferred word-vehicles akin to opting to speak psychology in Russian rather than in English, such disputes are simply non-cognitive quarrels over personal tastes and have no legitimate place in psychological science. But of course there is much more than mere value discrepancies at issue in such oppositions as behaviour theory *versus* phenomenology, trait-theoretical *versus* dynamic approaches to per-

[21] This is not a charge of complacency. There has never been a time when more ferment and probing has vitalized philosophy than in contemporary movements. But no matter how bustling the site, a construction in progress is not the same as an inhabitable edifice.

[22] (Ed.) Part III of the paper is not included here.

sonality, S-R *versus* cognitive interpretations of complex human behaviour, etc., even if it is often hard to make out what the significant points of divergence actually are. Here, then, is a major area for metapsychological research: To analyse various psychological perspectives or "programmatic" theories for how each nuance of expression, pattern of data interpretation, and whatever else may characterize the perspective's style makes a difference for what a person who adopts this perspective believes about psychology, or what it *enables* him to believe, in order that each such detail may be closely examined in naked isolation for its own particular intellectual merits. (Acceptance or rejection of a particular point of view inevitably carries with it adoption of numerous covert beliefs or belief-biasing attitudes to which we would not accede were we to recognize them for what they are, and the more holistic our allegiance or opposition to a given perspective, the more our thinking is corrupted—as judged by our own intellectual standards could we but bring them to bear—by such invisible cognitive demons.) I shall not here attempt to inventory the different *ways* in which a theoretical perspective can have import for what one believes even when the perspective makes no explicit factual commitments—though it is worth passing mention that a great deal of the content in perspectival theory is actually metatheoretical, and that an enormous share of past and present psychological controversy (e.g. the pros and cons of the behaviouristic approach) has been metapsychological conflict rather than disagreement over psychology proper.[23] I would, however, like to give two examples of how acceptance or rejection of a particular manner of speaking (verbal perspective) can carry powerful hidden commitments of which we are seldom aware.

(i) Learning theorists frequently hesitate to use the terms "elicit" and "evoke", to describe the relation between a stimulus S and its empirically correlated response R, on grounds that this implies too severe a reflex-type connection between S and R. The preferred locution for describing the effect of a discriminative stimulus on operant behaviour or (for some verbal learning theorists and virtually any psychologist whose perspective is dominantly "cognitive") the human emission of verbal responses to verbal stimuli is to say that S "sets the occasion for," "selects," or is a "cue" for R. There will probably be little disagreement that saying "S elicits (evokes) R" does, in fact, suggest that the $S \to R$ relation is akin to a "reflex"; that it has connotations which can be brought out more vividly by speaking of S as jerking, pushing or forcing R out of the organism; and that the physical analogy which comes most readily to mind is that of a doorbell ringing when the button is pushed. Alternatively, to say that S is a cue (selects, sets the occasion) for R is intended to convey merely that presence or absence of S is correlated with the

[23]For example, when one perspective takes another to task for neglect of certain data or possibilities, or contends programmatically that its framework ideas can eventually be worked into a significant and true propositional theory, the thesis concerns the actual or potential achievements of a certain theoretical approach and is hence a metatheoretical proposition.

organism's propensity to emit R. But are there any genuine *cognitive* differences between "S elicits R" and "S is a cue for R"? The fact that one suggests analogies the other does not is inconclusive, for to say that A is analogous to B is only to imply that there is some respect in which A resembles B, and not until the point of resemblance is made explicit can we decide whether or not C is also analogous to B in this way. That is, if by "elicit" we mean whatever sort of relation it is that links response to stimulus in a *reflex*, while we also agree (as surely we must) that this relation is not *literally* the same as the connection between doorbell and pushbutton, nor is it a literal jerking, pushing or forcing, then it remains to be seen whether those aspects of jerkings and bell-ringings which make reflexes appear analogous to the latter may not also appear, perhaps more subtly, in cueings and occasion-settings as well. Or turning the point around, what there is about bell-ringings and physical jerks and pushes which does *not* have a suitable parallel in discriminated operants and verbal respondings may not have a proper counterpart in reflexive elicitations either, and should hence not be construed to be implied by "elicits." Now, the most conspicuous formal property of these physical analogies, and the one which surely dominates the thinking of most persons who deny that verbal association performances or operant respondings are like this, is the near-perfect correlation between antecedent (e.g., button push) and consequent (bell ring)—and as a matter of brute fact, this is *not* an essential or even frequent property of reflexes. The strength of a reflexive association $S \to R$ is maximal when the probability of R, given S, is 1.0, but nothing in reflex theory demands that all reflexes be near maximal strength, and conditioned reflexes in particular usually involve low response probabilities in early formation or late extinction. Hence there is nothing in the concept of "reflex," and thus neither in "elicitation" or "evocation," which *properly* connotes inflexible coupling of R to S.

Even so, there are more subtle meaning ingredients in "elicit," derived from its paradigmatic application to reflex action, which do not quite fit most behaviour-theoretical conceptions of operant (instrumental) responding. In brief, the most important of these are that (a) reflex theory makes no provision for motivational determinants of behaviour;[24] (b) the reflex-theoretical conception of "stimulus" and "response" envisions these paradigmatically as a pulse of sensory excitation and a pulse of reaction, respectively, to which the notion of "latency" is applicable,[25] whereas the paradigmatic discriminated operant is a response which occurs repetitiously throughout the discriminative stimulus's continuation (i.e. it is the presence of S, not its onset, which is primary); and (c) the paradigmatic responses

[24] Pavlov (1928, p. 152) observed that degree of food deprivation influences the strength of conditioned salivation reflexes, but he did not conceptualize this *as* a motivational phenomenon.

[25] Sustained reflexive reactions to steady stimuli, as in prolonged pupillary constriction under continuous bright light, may be subsumed under the pulse paradigm by construing them as limiting cases of a rapid sequence of pulses.

in reflexive elicitation are specific motor discharges, whereas in operant behaviour they are achievements (i.e., changes in the environment or the organism's relation to it which can usually be brought about by a variety of motor sequences). Thus a psychologist may justly withold the term "elicitation" from an S-R correlation if he thinks motivation is also significantly involved in R's emission, if he doubts that the stimulus action is primarily an onset phenomenon, or if the response is defined as an achievement. On the other hand, if one objects to "elicitation" only because of analogies to invariant succession in mechanical systems, he has not merely failed to distinguish between literal meaning and metaphor in behaviour theory, but has also implicitly ascribed to psychologists who *do* speak of elicitation views which in all probability they do not in fact hold.

(ii) While linguistic style is one feature which usually helps to distinguish one perspectival theory from another, one would like to think that preferences in grammatical phrasing make little difference for a theory's cognitive impact. But such is not the case, as attested by the remarkable increase in conceptual power which accompanies the shift from transitive to intransitive verb-forms. Consider, for example, the contrast between "John sees (perceives) the stopsign" and "John has a perception (percept) of the stopsign." A trivial difference? Far from it. The transitive verb regards perception as a primitive relation between the perceiver and the perceived, whereas the intransitive construction construes perception as mediated by a condition of the perceiver which is what stands more directly in the aboutness relation to the perceived. That is, "s perceives o" has the logical form

$$P_1(s, o),$$

whereas the logical form of "s has a percept of o" is [26]

$$(\exists\phi)[\phi(s) \cdot P_2(\phi, o)]$$

There is no incompatibility between these two formulations. A theory of perception which prefers intransitive constructions can always maintain that "s has a percept of o" is an *analysis* of "s perceives o," while a perspective which prefers transitive verbs concedes the analytic equivalence of the two forms so long as it is agreed—if it is—that when a person perceives something he does so in virtue of certain features of his momentary psychological condition which constitute his perceiving one thing rather than another. Yet a transitive perspective in perception theory is inherently focused upon the objects of perception and has no

[26] More precisely, "s has a percept of o" should perhaps be formalized as $(\exists x)[H(s, x) \cdot P(x) \cdot R(x, o)]$, in which H is whatever relation holds between a percept and the person who has it, P is the attribute of being a percept, and R is a referential relation. However, the nature of H is presumably exemplification, while P can be combined with R to constitute a relation P_2 of perceptual aboutness.

linguistic resources (except insofar as these are subsequently introduced by deliberate contrivance) with which to think about the *means* of perception, whereas an intransitive perspective has explicit concern for the mechanisms by which perception occurs, and even a rudimentary propositional theory thereof, built into its very grammar.

The difference between transitive and intransitive verbs is, of course, relevant to many other psychological issues besides perception. In fact, a great deal of commonsense psychology is couched in transitive forms whose replacement in our thinking with more articulated intransitive counterparts is requisite to getting on with technical progress in these areas. But unfortunately, intransitive constructions don't have the same subjective warmth and liveliness as do their transitive precursors. They seem to construe the person as a passive, psychologically demeaned and impoverished thing whose sole function is to provide a point of confluence for such impersonal entities as sounds, smells, twitches, and the like. And a psychological perspective in which human beings appear as colorless containers for what happens to them, rather than the active agents we know ourselves to be, readily provokes rejection and hostility from critics who interpret such a perspective as an attack on human individuality, responsibility and personal worth. Actually, this reaction is without cognitive foundation. The organism's apparent psychological emasculation by intransitive verb constructions is simply an instance of the holistic delusion that analysis is denial of the thing analysed. Yet wherever in past and present psychology the concept of "self" has been esteemed, the noncognitive feeling tones of transitive grammar have been an impediment to the growth of more penetrating theory, while even in areas where there is no fear of devaluating humanity, seduction by transitive grammatical habits continues to procreate conceptually crippled theory.[27]

The preceding two examples are small illustrations of how points of tension between divergent theoretical perspectives can be amorphous fusions of emotion-charged irrelevancies with genuine cognitive issues which mayor may not sustain serious interperspectival hostilities or intraperspectival uncertainty when they are examined openly. But metatheoretically problematic quasi-conflicts can be found at all levels of the perspectival/propositional continuum. Even where assertions are most definite in their substantive commitments, it is still not at all uncommon for theories which apparently contradict one another to be in fact perfectly compatible or even to lend each other active support. Some of the most blatant instances of this occur when different patterns of organizing abstractions (theory$_{\mathrm{B3}}$) or idealizations (theory$_{\mathrm{B4}}$) are contrasted. For example, alternative non-linearly related choices of scale for a given variable lead to different statistical conclusions about its expected value in a given population; while idealized equations of grossly

[27]E.g., hypotheses about subjects' transitive "use" of mediated associations in verbal behaviour, aptly characterized by Mandler (1963, p. 247) as "homunculus-like."

different algebraic form can be virtually indistinguishable over the region of the data to which they are fitted, or, even if the idealizations themselves differ significantly, the residual variance in the data may be no worse under the one than under the other.

Thus two summary statements of what, on the whole, a given body of data is like, especially in regard to the patterns of relationships therein, can appear to disagree even though both are in fact equally true. Or at least they *may* both be true. It is also quite possible that different ways to summarize the same extant data take the form of generalizations which become irreconcilable when extrapolated to predict what would happen if the range of the independent variables were extended, the background constancies shifted, or extraneous variables more tightly controlled. Since assertions of empirical regularity (i.e., observation-language theories) are universally vague about the boundary conditions under which they are considered to hold, it is a job for metatheoretical analysis to decipher just where actual disagreement begins in different accounts of the same body of data, and what its nature may be.

To be sure, in these days of relatively enlightened quantitative methodology, few research psychologists are disposed to quarrel over mildly discrepant idealizations treated merely as summaries of extant data, especially when these are described as "models" with all the bland tolerance for inaccuracy and untoward implications this term encourages. Even so, one tends to feel that alternative "models" of the same data are working at cross-purposes in that if each *were* elevated to the status of indicative theory, then they *would* be conflicting hypotheses about the events and regularities underlying these observations. I think it is fair to say that for the most part, the really embattled conflicts about psychological reality concern theoretical entities in sense B1, and that theorists are often disposed to regard non-synonymous B1-type theories over the same empirical domain as logically incompatible. But in fact, there are good abstract-metatheoretical reasons for thinking that two theories about unobservables can be in factual conflict only if they entail an observational disagreement, no matter how contradictory their theoretical$_{\mathrm{B1}}$ postulates or theorems may appear (Rozeboom, 1962), and even quite possibly that they must make essentially the *same* theoretical$_{\mathrm{B1}}$ commitments to the extent that they are warranted by the same body of data. Thus it may well be that two propositional theories (or the more propositional aspects of two theoretical perspectives) which are thought by their proponents to be at loggerheads actually say pretty much the same thing, though not necessarily so in any *obvious* way.

For example, by far and away the most celebrated controversy in psychology's history has been the running battle between behavioural (S-R) and mentalistic (cognitive) theories. We all know of the agony and outrage with which so many mentalistically oriented psychologists have excoriated S-R theory as a denial of man's inner existence, and how, conversely, S-R accounts of overt actions often seem to be thought by their partisans to render mentalistic interpretations otiose. Yet so long as behaviour theories are not couched entirely in their observation language (and none has ever been, not even Skinner's), then it is altogether possible that the theoretical$_{\mathrm{B1}}$ terms introduced by a behavioural hypothesis to account for certain overt regularities *refer to* (designate, are about) the very same internal entities that we know more familiarly in mentalistic terms, even though the behaviour-theoretical terms do not have the same *meanings* as the mentalistic concepts with which they are co-referential. For example, if an S-R theory of motivation finds it necessary to postulate an "intervening variable"[28] D_T which varies as a function of water deprivation and in turn influences, *inter alia*, the intensity of drinking behaviour, D_T may in fact be the very same variable that we introspectively identify as "thirst" even though the meaning given to D_T by its postulated connections with peripheral Ss and Rs does not logically require that D_T be identified with subjective thirst. Although we are still a long, long way from it today, there is no reason at all why a suitably sophisticated behaviour theory whose theoretical$_{\mathrm{B1}}$ terms have only stimulus- and response-conferred meanings cannot nonetheless be *about* all the rich, warm inner experiences that we are now able to address only through the undisciplined conceptual resources of phenomenology and commonsense mentalism.[29]

Moreover, even when the individual theoretical$_{\mathrm{B1}}$ terms and propositions of two *prima facie* competing theories do not have a simple one-to-one correspondence in factual reference, they may still be factually equivalent on a larger scale. For example, a good deal of printer's ink has been spilled in inferential factor analysis over how best to solve for common factors, the chief dimensions of controversy having concerned orthogonal *versus* oblique solutions and simple-structure versus hierarchical factor patterns. Now, extraction of common factor space by estimating data-variable communalities is an act of ampliative (i.e., not deductive) inference to underlying source variables (Rozeboom, 1966a, p. 252ff) which is certainly a worthy subject for controversy. But for the most part, the big disputes in factor theory have concerned not the solution for common-factor *space*, but where the source variables lie *within* that space—i.e., communality estimation has generally been treated as a negotiable detail while the battle lines have been drawn over rotation to terminal axes. But if $F_1, \ldots, F_n$ and $F'_1, \ldots F'_n$ are different choices of axes for the same common-factor space, then each F_i is an

[28]Really a hypothetical construct, though since psychologists have refused to use the term "intervening variable" in the sense given to it by MacCorquodale and Meehl (1948), I will not fight the current here.

[29]This is essentially what is known in philosophy as the "identity" solution to the mind-body problem (see Feigl, 1958), except that I am here focusing upon the possible co-reference of mental concepts with behavioral rather than neurophysiological constructs.

exact linear combination of the F'_j and conversely, and any hypothesis about the nature and functioning of one or more of the inferred source variables $F_1, \ldots, F_n$ is true if and only if a corresponding (though quite possibly more complicated) statement about the F'_j is also true. At the root of the rotation controversy is the obscure but compelling notion that since inferential factoring is a hypothesis about underlying (non-observational) reality, it must also be assumed that some of the dimensions (variables) in common-factor space are *more* real than others. This is the very same intuition which causes us to feel, say, that a person's weight-in-pounds and height-in-inches have an existential solidity not possessed by measures such as "whyght," defined as weight-in-pounds plus height-in-inches, analytically abstractable from the former. I am by no means convinced that this feeling is wholly without foundation, but it concerns exceptionally esoteric philosophical issues on which, to my knowledge, no factor analyst has ever *intended* to take a stand. Thus apart from presumably unwanted overtones, it would seem that different styles of inferential factoring disagree more over how the total inference about source variables is to be partitioned conceptually than over what is actually hypothesized to be the case about underlying reality. Or more precisely, while there may well be genuine cognitive issues involved in the rotation problem, it is still highly obscure just what they are.

To summarize the present point, then, if *prima facie* competing theories are pruned of their unintended presumptions and ingenuous metatheoretic denials of all rival positions, it may turn out that they say essentially the same thing or at least supplement, rather than contradict, each other albeit through the media of different conceptual frameworks. Determining the extent to which this is so is hence another task at which technically skilled metatheoretical analysis can be of value. Or is this of value? If we but leave intertheoretic squabbles to those who have nothing better to do, will it not suffice for each theory to take care of its own growth and revision? No; because this idyll of deep down intertheoretical accord I have been suggesting is only a possibility, not a to-be-expected actuality. Competing theories often contain genuine disagreements along with the spurious ones, and there is probably no faster way for theories to evolve than for one to be pitted against another on points of real issue. Or one theory may have developed a concept, solved a problem or recognized a phenomenon which would greatly benefit another were it translated into the other's terms. The encompassing point is that the ways in which contraposed theories supplement, contradict or concur with one another are often very different from the ways in which they *seem* to do so, and it is hardly conceivable that a more veridical perception of their interrelations would not be of immense value for development and application of the theories themselves.

The logical structure of theories

The tasks which I have commended to metapsychology so far may be described generically as an explication of what the cognitive ingredients of a given theory actually are. And arch as this phrase, "explicating cognitive ingredients," may sound, there is a solidly operational way for metatheoretical practice to realize its vision, namely, by laying bare the *logical structure* of the theory's concepts and tenets. As is true of any technical skill, the actual procedures and products of such analysis can be properly appreciated only through observing it at work in serious applications, and since part III of this paper[30] attempts to communicate a modicum of this appreciation through deeds, there is little to be gained at this point by loitering over words. But insomuch as *Problems of Logical Structure* is one of the two primary categories into which metatheoretical issues fall, I would like to say one or two things about what, abstractly, is involved here.

In brief, the logical *structure* of a theory is the *pattern* by which its concepts and assertions, if any, hang together. The following semi-technical definition will clarify this somewhat. Let $\mathbf{E}^T = \{E_i^T\}$ be the set of all meaningful linguistic expressions—individual terms, sentences, compound predicates, or whatever—contained by a theory T; let $\mathbf{S}^T = \{E_j^T\}$ be the set of all meaningful sentences which can be generated by concatenating expressions in $\mathbf{E}^T$; and let $\mathbf{\Sigma} = \{\mathbf{E}^k\}$ be the set of all subsets of all meaningful linguistic expressions (not just those found in T) such that each $\mathbf{E}^k$ in $\mathbf{\Sigma}$ also satisfies the following condition: There exists a mapping ϕ^k of expressions in $\mathbf{E}^T$ into expressions in $\mathbf{E}^k$, and derivatively of all concatenations of expressions from $\mathbf{E}^T$ into corresponding concatenations of expressions from $\mathbf{E}^k$, in such fashion that for any S_i^T and S_j^T in $\mathbf{S}^T$, if S_i^T analytically entails S_j^T, then $\phi^k(S_i^T)$ and $\phi^k(S_j^T)$ are meaningful sentences such that $\phi^k(S_i^T)$ analytically entails $\phi^k(S_j^T)$. Then the logical structure of system $\mathbf{E}^T$ (i.e., theory T) is whatever is common to all systems in $\mathbf{\Sigma}$ i.e., it is whatever is described by a set of rules which tell how to change the meanings of the symbols used by theory T without changing the entailment-relationships among sentences constructed from these symbols.[31] For an ideally precise and explicit theory, this is essentially the

[30](Ed.) Not included here since the "explication of cognitive ingredients" of theories is also a major focus of some later papers in this volume, notably "Problems in the psycho-philosophy of knowledge", "Good science is Abductive", and "The problematic importance of hypotheses".

[31]While this definition of "logical structure" is not completely explicit (especially in regard to sentence formation through concatenation), it has been rather carefully worded to avoid any presupposition that theory T has been formalized. To search out logical structure in a living language teeming with synonyms (meaning equivalences not revealed by symbol identity), ellipses and suppressed premises (sources of meaning relations among statements not revealed by their overt grammatical construction), and the like, we need a definition of structure which does not require any initial assistance from the physical properties of the language's sign-vehicles. The entailment cited here is not formal but analytic (semantic), so that a sentence S_i can entail sentence S_j because the meanings of S_i and S_j require that S_j be true if S_i is, even though typographically S_i and S_j may have no formal connection. Weaker forms of implication (i.e., inductive support) are also relevant to disclosure of logical structure, but presumably any structural property which

syntactic structure abstracted in the logician's concept of "uninterpreted" theory (A2), except that it is the *spirit* of formalization which is vital to applied metatheory, not the typographic accoutrements of logistical or mathematical notation. As it is, the primary goals which mathematicians and logicians have sought to achieve through formalization, namely, consistency, deductive impeccability, and parsimony of postulates and logical machinery in axiomatic development of criterion theorems already well established by informal argument, have only secondary importance for a science like psychology wherein beliefs remain in a flux of emendation and overthrow even where they can be found at all. Deductive accuracy is unquestionably desirable (and rarer than we like to think) in science, as is being able to trace out the implications of a set of premises beyond the point where technically unamplified reason falters. But extant mathematical or logical formalisms which enhance deductive power do so precisely because they suppress all logical structure which is not directly germane to the particular inferential algorithms there exploited, not to mention the distortion which may have to be introduced into the original theory in order to satisfy the formal system's presuppositions. With our present still-primitive awareness of conceptual structure, an approach to theory construction which seeks to promote deductive rigour by *replacing* the half-inarticulate but semantically intricate verbal ventures of a living science with a computer-programmable formal calculus will inevitably freeze out most of the subtle complexities to which the theory owes its perspectival appeal and capacity to evolve.

The most important role of formalization in scientific theory lies not in constructive applications wherein theorems are derived algorithmically, or new belief-systems are created wholesale by tacking meanings onto a pre-existent abstract calculus by means of correspondence rules, but in *re*constructive analysis which brings to light the theory's implicit structure. A given array of semantically linked propositions can usually be formalized in many different ways at different levels of complexity, and their finer logical texture is almost certain to escape us at first. But by formalizing as much of the theory's structure as we are now able to perceive, we construct templates which, when held against the living reality of the theory's usage, throw into bold relief what the theory has in it *beyond* what we have captured so far. For example, suppose that a certain theory advances a natural language argument in which conclusion q is held to follow from premise p, but that when p and q are replaced by their *prima facie* formalized equivalents ϕ and ψ, respectively, ψ is not a deductive consequence of ϕ. Three main alternatives now present themselves. The first is to accuse the theory of a logical fallacy, and this would indeed be legitimate if, after the demonstration that ϕ does not formally entail ψ, q no longer seems to follow from p. (In this way it might be discovered, e.g., that the earlier informal argument implicitly assumed a

mediates weak implication will also be manifested elsewhere in strict entailment.

second premise r which is needed to supplement p before q follows.) Secondly we may decide that for any two propositions respectively formalizable by structures ϕ and ψ, the former still counts as good *evidence* for the latter even though the relationship between them is not one of deductive implication. In this case the formalization has isolated a pattern of inductive inference which may or may not have been recognized previously (and if not, we have turned up a new datum for inductive logic to assimilate), but contributes no new powers to a theory which has been using this pattern of inference all along. Finally and most significantly, the natural-language proposition q may still seem to follow from natural-language premise p even after we have recognized that an argument of form "If ϕ, then ψ" has no generic merit. In this case, we must conclude that formalizing p as ϕ and q as ψ fails to capture those logical connections between p and q in virtue of which the former supports the latter, and we are accordingly directed to reanalyse p and q more deeply in search of why p is evidence for q, thereby disclosing ingredients in p and q which we had previously overlooked.

Let me give a more concrete illustration of this general point, for it is an important one. Suppose that I say, "John is infatuated with Marsha, but doesn't want to marry her." How is this to be translated into logistical notation? If we write "$I(x, y)$" for the predicate "x is infatuated with y," "$M(x, y)$" for the predicate "x wants to marry y," "a" for "John" and "b" for "Marsha," my statement about John's feelings for Marsha analyses most obviously as

$$I(a, b) \cdot \sim M(a, b).$$

However, if we translate this formalization back into everyday English, we have "John is infatuated with Marsha *and* it is not the case that he wants to marry her." That is, "but" passes over into "and" in the formalization. Now, as soon as we recognize this it becomes apparent that the original statement conveyed more than just the co-occurrence of two relations, I and $\sim M$, holding between entities a and b. The conjunctive "but" carries an overtone of surprise which softly implies, without actually asserting it, that $\sim M(a, b)$ is contrary to expectation given $I(a, b)$. The expectation presumably rests upon some background regularity connecting I and M, though it is not at all clear what relationship is envisioned (e.g. "Almost everyone wants to marry persons with whom they are infatuated" and "John usually wants to marry girls with whom he is infatuated" both support the "but"-surprisal in this case) or how strong it is. In any event, if I am trying to be maximally explicit in what I have to say about John and Marsha, I now have two choices: (*i*) I can abandon "but" for "and" and claim *merely* that John is infatuated with Marsha and doesn't want to marry her, or (*ii*) I can state this and augment it with some generalization linking infatuation and matrimonial desires in virtue of which this particular conjunction is atypical. What I am *not* entitled to do is to remain pat with my original "but"-statement, for this has now been shown to contain a lawlike allegation which has not been expressed in a form amenable to cognitive evaluation. (Of course, no one can *compel* me to make any revision of my original statement, but if I am not willing to do so, then my continued use of "but" in this context is intellectually dishonest.)

And what does the logistical formula "$I(a,b) \cdot \sim M(a,b)$" contribute here? Its virtue lies not in a dehumanizing reduction of John's feelings for Marsha to an inflexible concatenation of meaningless marks in an abstract calculus, but in its precise expression of a plausible paraphrase for my original statement, a paraphrase which I am challenged either to accept or to compare with the original and make explicit the surplus meaning so disclosed. Notice that the full value of the formalization emerges only when we translate it *back* into a semantically familiar form. In principle, we could carry out this comparison without ever leaving ordinary English. The advantage of the logistic mediation is merely (though this is considerably more than just a "mere") that this compels our analytic paraphrasing to emphasize those grammatical forms which are most clearly understood in technical logic. As it is, the present example also serves to illustrate why it is dangerous (or would be were there any appreciable likelihood of this step's being taken) to *work* with one's concepts and hypotheses in logistic format. The predicate "x wants to marry y" contains immensely more logical structure than is represented by the dyadic form "$M(x,y)$". For one, the tensed verb in the original predicate makes a temporal reference which is altogether suppressed in "$M(x,y)$". More important, *wanting* w is most penetratingly analysed as having a want related referentially to w (cf. earlier comments on transitive versus intransitive verbs), and the grammar of want-descriptions is on the same level of logical complexity as descriptions of memories, percepts, beliefs and the like[32]—none of which is remotely suggested in the formalization of marriage-wanting as a dyadic relation lacking internal grain. The fine structure of want-concepts is essential to much of the intuitive reasoning we do with them, and can eventually be made explicit by successive attempts to formalize such arguments. But here as in all such cases the original concept or statement is still the onion off which a particular formalization peels only certain restricted layers of structure.

The truth/veridicality of theories

Analysis of logical structure answers questions about what, cognitively, a theory has in it. (It's not the whole answer, of course, since something must also be said

[32]In primary occurrences of the grammatical form "Person s wants (remembers, perceives, believes) ____," the blank is filled with an expression whose logical structure is that of a complete sentence, though components of this are frequently suppressed in practice. Thus "x wants *to marry* y" is more accurately written "x wants *that* x *will marry* y." It may be noted, incidentally, that "John doesn't want to marry Marsha" is not properly formalized as $\sim M(a,b)$ at all if what is intended by the original is not a simple lack of marriage desire in John for Marsha, but rather, John's wanting *that John will not marry Marsha.*

about meaning. But meaning becomes to a surprising degree superfluous once logical structure has been disclosed. E.g., in a science whose concepts were so exhaustively formalized that *all* their meaning relations were syntactically explicit, meanings would make a difference only when raw observations call forth datum assertions.) Moreover, successful explication of a theory's logical structure to any depth desired requires no more than persistent application of concept-analytic routines, even if proficiency in these skills and indeed even appreciation for the ends they serve is still conspicuously lacking in contemporary psychology. But there is a second cluster of metatheoretical problems, just as fundamental as the first, for which no simple resolution routines exist, namely, *Problems of Veridicality.* While straightforward declarative assertions are not nearly so prevalent in psychological theory as traditional metascience would lead one to expect, it is surely the ultimate *goal* of psychology to arrive at true statements about psychological reality. A perspective remains unfulfilled so long as it is merely perspectival, while idealizations (B4) and models (B5) are truth-claims playing it cool. But when is a theory true (correct, veridical), anyway?

Now, Truth in all its philosophic splendour is far too involved and ponderous a concept to be invoked here with useful issue. Fortunately the problems of veridicality which confront applied metatheory can be reworded to avoid it. To ask whether or not a proposition p is *true* is for all practical purposes equivalent to asking whether or not we should believe p. Thus the root problem of veridicality may be rephrased as *when is a theory T credible*? Or better, since credibility is a matter of degree, to what *extent* is T credible (plausible, warranted, tenable, believable, acceptable), and why? It can, I think, be agreed without much argument that research psychologists are to a man profoundly committed to weighing and altering the credibilities of various psychological "theories" in one sense or another of this term. This is, after all, the definitive process of science: We start with beliefs at the highest level of assurance we can achieve, i.e. datum-convictions, and move from there to (hopefully) appropriate degrees of belief in certain other propositions about which we would like to be certain; and all our most elaborate techniques of experimental design and data collection are simply methods for generating datum-beliefs which drive the tenability of these questioned propositions toward one extreme or another. Since research science is so extensively engaged in the *practice* of assaying theoretical credibility, it is meet that metatheory identify and, ultimately, pass normative judgment on the forms which govern these assessments—except that such norms best arise not by one man's fiat of how another must conduct his thinking, but through so detailed an explication of the patterns by which a theorist *does* transfer credibility from one proposition to another that the theorist himself, given this heightened awareness of his own inferential methodology, can judge whether this procedure still appeals to his reason in quite the same way as before, and if not, what purifications of method now support his conclusions more strongly. Mankind is still very much in the throes of evolving its powers of non-demonstrative inference, and while this has reached its most powerful development in the professionally disciplined construction and evaluation of scientific theories, I shall now show that what makes a theory credible is considerably more problematic than has generally been recognized, and that the region of uncertainty has major operational import for practical theory building.

Extant metatheoretical interpretations of how scientific theories wax and wane in their cognitive respectability may be sorted rather neatly into three primary perspectives, *inductivist*, *hypothetico-deductive*, and *omnitheoretic*. Inductivist views hold that there exist specific patterns of nondemonstrative inference according to which properly organized datum-beliefs project credibility beyond themselves. With respect to Reichenbach's celebrated distinction between the "context of discovery" and the "context of justification," the inductivist outlook substantially equates the two—the data are thought to call their inductive implications to mind as well as to endow them with evidential support. Although the vast majority of practicing scientists probably subscribe to some version of the inductivist position, it is regarded as rather naive and hopelessly inadequate by many contemporary philosophers of science. Instead, most abstract metatheoreticians have until very recently favoured the hypothetico-deductive view. This holds that how a theory originates in imagination is inexplicable by any pattern of formal logic and is, in fact, irrelevant—all that matters for the theory's epistemic stature is subsequent verification or refutation of the theory's testable consequences. And finally, objecting to both inductivist and hypothetico-deductive positions on grounds that they presuppose a theory-free base of observations, is the nascent omnitheoretic development. None of these views are wholly satisfactory in their extreme versions, but I shall argue that the inductivist position is at least *basically* correct in a way that the other two are not.

The omnitheoretic approach Actually, the omnitheoretic perspective has not as yet advanced any firm thesis about the determinants of a theory's credibility, though this movement's most dedicated spokesman, Paul Feyerabend, has been explicit in his judgment that traditional hypothetico-deductive and, presumably, inductivist accounts are untenable because they assume datum beliefs to be independent of the theories for or against which they are cited as evidence (Feyerabend, 1963). (Hanson, 1958) and (Kuhn, 1962, Ch. 10) have proposed that the conceptual onset of a theory is essentially a perceptual phenomenon—i.e. that scientists do not *infer* a theory from the data, but somehow begin to *see* their data *as* an embodiment of the theory, while since these new datum-perceptions contain the theory as a presupposition, they are just not the same as the datum-perceptions which would have been made had the theory not been accepted (perceived therein). *But this is simply to claim that rational judgment does not enter into acceptance* *or rejection of theories at all*. For rational judgment is above all a process of inference from one proposition to another, and if the theory's acceptance or denial is given in our very basis for inference, then there is nothing further to be done—the theory's status has been settled for us before we can even start to pass judgment on it. In particular, if datum-proposition D entails (presupposes) theory T, then D can imply $\sim T$ only insofar as D is self-contradictory, in which event D is scuttled as a datum-belief and T is no more impugned than before. (Kuhn argues that scientific revolutions replace old theories with new ones when crises and tensions—i.e., data which do not fit comfortably—become too much to bear. But under the omnitheoretic view tensions should not be able to arise in the first place.) The only way for a theory to be discarded would be for us to undergo a spontaneous (i.e. nonrational) shift of perception. Further, the mere discarding of a theory is not the same as *disconfirming* it, any more than e.g. putting aside our present system of weights and measures in favour of the metric system would be a denial of the numerical readings we took in the old units. Even if our new perceptual theory has the appearance of contradicting the old one, the omnitheoretic contention that observation-language terms change their meanings with changes in theory (cf. Feyerabend's rejection of "meaning invariance") implies that there need be no logical inconsistency between the sentence-complex T when it asserts an accepted theory and the sentence-complex $\sim T$ asserted when the theory previously expressed by T has been abandoned. Hence, not only are we unable to disconfirm accepted theories, we cannot even know that they were wrong after we have relinquished them.

The undeniable fact that scientists and laymen *do* make inferences from datum-beliefs, while any proposition to which inference is made can be construed as a "theory," shows that the omnitheoretic thesis is absurd unless severely attenuated in some way. One mandatory retreat is that datum-beliefs must be conceded neutrality with respect to at least *some* of the theories to which they are cognitively relevant. In the modified omnitheoretic view, then, the datum-perceptions D of a given scientist at a given moment carry certain theoretical commitments T_D, but any theory T whose negation is compatible with T_D can be argued for or against on the basis of D by whatever standards of inference are acceptable on non-omnitheoretic grounds. Given this reformulation, the omnitheoretic thesis becomes generically undeniable (as discussed under BI, perceptual judgments in data collection are never incorrigible in the toughest philosophic sense but always incorporate hypotheses of one sort or another), but the extent to which it advances our metatheoretic understanding of scientific knowledge is highly moot. It was pointed out earlier that when scientific controversy begins to involve questions which differentially bias the datum-perceptions (or at least the verbalised datum-statements) of an area's researchers according to their individual theoretical preconvictions, as signalled by their inability to reach consensus even on datum-statements, strong pressures arise toward development of methods for data collection and description which are, in fact, neutral with respect to the theoretical question at issue, though not necessarily so regarding presuppositions which have not as yet become controversial. Such neutrality is not easily achieved, but its pursuit is one of the primary *technical* phases of science (and hence easily obscured or left incomprehensible in commonsense descriptions of scientific activities) and the history and contemporary practice of the various sciences makes abundantly clear that through sustained effort and professional skill, the observation-language impartiality requisite to *reasoned* judgements on particular points of theoretical issue can be adequately approximated in operational reality.[33] While the omnitheoretic thesis usefully reminds us that a naive empiricism which considers theoretical neutrality in datum-beliefs to be the natural state of primitive perception is badly misguided, and that one important mode of scientific progress is the liberation of datum-beliefs from presuppositions which have become problematic, it does nothing to illuminate how a science's observational premises support or disconfirm theories which are *not* presupposed by the science's datum basis.

The hypothetico-deductive approach We next turn to the hypothetico-deductive account of theory development. While this is the perspective which has dominated the past several decades of metascientific theory, I shall demonstrate that it fails abjectly to provide any acceptable standards for the confirmation of theory by data. According to the hypothetico-deductive thesis, how a theory T first arises—i.e., what makes its proponents first think of it—is neither relevant to its credibility nor (at least in Popper's extreme views) can be accounted for by any formal inference patterns. Once T has emerged into conscious consideration, however, it can be supported or disconfirmed through testing of its observational consequences. Specifically, let C be some observation-language proposition such that C is a logical consequence of theory T,[34] while the truth of C has not as yet been settled.[35] Then the credibility of T is enhanced if C becomes verified while, of course, T is disproved if C is shown to be false. So far, the hypothetico-deductive thesis is unassailable. Under any acceptable idealization of rational inference, if T entails C, then confirming C even probabilistically likewise increases the credibility of

[33] I submit that if the many provocative examples of scientific "revolutions" cited by Kuhn (1962) are analysed closely, it will be seen that they either in fact did proceed or, with a little more methodological acumen, readily could have proceeded by argument based upon observational assertions mutually acceptable to all parties concerned.

[34] The hypothetico-deductive thesis readily extends to cases where T implies C only probabilistically. But deductive consequences are simplest to discuss and quite suffice to make my present point about this position's inadequacy.

[35] Discussion of the manner in which theory T derives support from a consequence C whose truth was known prior to its derivation from T involves some technical niceties which since they add nothing of importance to my main argument, I shall avoid by assuming that C becomes verified only after its deduction from T.

T. But the hypothetico-deductive account also *stops* at this point and, by thus ignoring further analysis of T, invites the inference that verification of T's consequence C increases the credibility of T *throughout*—i.e., of all of T's conjunctive components.

Acceptance of this invitation, however, would bring total disaster to the scientific enterprise. For suppose that T is a theory which entails testable consequence C, while T^* is the conjunction of T with some arbitrary hypothesis H (i.e. $T^* =_{\text{def}} T \cdot H$). Then T^* also entails C, and verification of C increases the credibility of both T and T^*. But if verification of one of a theory's consequences confirms the theory throughout, as the hypothetico-deductive thesis tacitly urges, *then confirmation of T^* by verification of C strengthens the credibility of T^*'s component H even though H has no relevance for C.* For example, let T be a theory of celestial mechanics which has as a consequence that the sun will rise tomorrow morning, while T^* is T conjoined with the hypothesis that my wife is unfaithful. Then by the hypothetico-deductive argument, the sun's rising tomorrow morning should confirm my suspicion of my wife's infidelity. By this line of reasoning, the most bizarre hypothesis can be given repeated empirical support by attaching it to a legitimate theory which has many verifiable consequences.

Now, while hypothetico-deductive holism *seems* to imply that affirmation of a theory's consequences indiscriminately confirms all the theory's components, the position does not actually *assert* this, nor could it do so without becoming trapped in a fundamental inconsistency. For suppose that T entails C while $T^* =_{\text{def}} T \cdot H$ and $T^{**} =_{\text{def}} T \cdot \sim H$. Then verification of C confirms both T^* and T^{**}, and if this support were to penetrate to all conjunctive components of the theories confirmed we would have a simultaneous increase in the credibilities of both H and $\sim H$, in violation of the cardinal metatheoretic rule that an increase in the credibility of any proposition must be accompanied by a decrease in the credibility of its negation. The hypothetico-deductive thesis must therefore concur with the following negative principle: *If C_1 and C_2 are both unverified deductive consequences of theory T, subsequent verification of C_1 necessarily increases the credibility of T as a whole, but need not increase—and in fact may even decrease—the credibility of C_2.* In particular, this holds when C_2 is one of the postulates whose conjunction composes T. We shall look more deeply into this situation in a moment. The immediate point is that the hypothetico-deductive argument provides no basis for distinguishing those of a theory's components which have evidential backing from those which are entirely arbitrary or are perhaps even discredited by the data. It does not help to argue that if we continue to test T's consequences long enough (say by a series of "crucial experiments"), we should eventually find one which is false if anything in T is unsound. When judging a theory's credibility, we are in possession only of these datum-beliefs which *have been* acquired, never those which *will be*, and to have any practical significance a theory of scientific inference must address itself to what can reasonably be inferred from the datum-beliefs which we in fact hold now.[36]

The analytic problem of theory confirmation, which hypothetico-deductive holism simply ignores, is thus: *If theory T is equivalent to the conjunction of postulates $P_1, \ldots, P_n$ while C is a logical consequence of T, in what way does verification of C affect the credibility of postulate $P_i (i = 1, \ldots, n)$?* And our discussion of arbitrary additions to an otherwise sound theory, whereby one might try to confirm his paranoid suspicions of his wife's infidelity by observing the sunrise, makes evident the following descriptive/normative[37] principle: *If theory T can be analysed as the conjunction of postulates $P_1, \ldots, P_n$, while C is a logical consequence of $P_1, \ldots, P_{n-1}$, then the fact that T logically entails C is irrelevant to whether or not verification of C confirms P_n.* That is, verification of a theory's consequences can be trusted to confirm at most those components of the theory which actually contribute to the derivation of these consequences.

But can we then salvage the hypothetico-deductive argument by proposing that verification of T's consequence C at least confirms those theory components which are needed to deduce C from T? Even this won't help, as shown by the following partition. Let C be a consequence, or the conjunction of a set of consequences, of theory T (i.e., $\vdash T \supset C$) while R is defined

$$R =_{\text{def}} \sim (C \cdot \sim T)$$

(i.e., R denies that T is false while C is true.) Then it is easily proved that (i) T is equivalent to the conjunction of C and R (i.e., $\vdash T \equiv C \cdot R$); (ii) unless C is logically true, C *is not* a consequence of R (i.e., $\vdash R \supset C$ if and only if $\vdash C$); and (iii) R is the weakest assertion with which C can be supplemented to yield T (i.e., for any sentence S, if $\vdash C \cdot S \equiv T$, then $\vdash S \supset R$).[38] This says that given any consequence C of any theory T, T can be expressed as the conjunction of C with a residual R which is what remains of T when C is factored out. Then confirmation of T by verification of C is a trivial outcome of the fact that this

[36] It is noteworthy that while Popper's version of the hypothetico-deductive position advocates bold, highly *confirmable* (corroborable) theories, where the greater the theory's logical content (i.e. factual commitments) the more confirm*able* (corrobor*able*) it is, the degree to which a theory is actually confirm*ed* (corroborat*ed*) at any given stage of testing is inversely related to its logical content even in Popper's account. Specifically, suppose that $T_1, T_2, \ldots, T_i, \ldots$ is a series of theories such that T_1 entails C and for each i, T_{i+1} entails but is not entailed by T_i. Then by either of the two confirmation measures proposed by Popper (1959, Appendix ix), confirmation of T_i by verification of C is a decreasing function of i.

[37] Descriptive in that it is a pattern by which we *do* find ourselves reasoning; normative in that we also feel that it is *right* for our beliefs to be so patterned.

[38] (i) and (ii) are obvious consequences of the fact that R is equivalent to $C \supset T$. (iii) follows from the fact that $S \supset R$ is equivalent to $\sim (S \cdot C \cdot \sim T)$ while the latter is equivalent to the tautology $\sim (T \cdot \sim T)$ so long as $\vdash C \cdot S \equiv T$.

confirms T's component C.[39] The scientifically significant question, however, is not whether verification of C confirms C, but whether it lends support to anything in T over and above C. (For example, if T is the hypothesis that all αs are βs while C is the datum that all αs observed so far have been βs, the problem of statistical inference is not whether C strengthens the credibility of T but whether it increases the plausibility that αs yet unobserved are also βs.) Doubly embarrassing to the hypothetico-deductive thesis is that in the $C \cdot R$ factoring of T, R is *discredited* by C under most any plausible formalization of credibility relationships.[40] Hence not only can a theory T always be conjunctively partitioned in such fashion that its only component needed to deduce its consequence C is C itself, but the credibility of T's remainder under this partition receives no support whatsoever from C's verification and is in fact generally disconfirmed by it. To be sure, this does not imply that verifying T's consequence C never confirms anything in T beyond C itself, but it does make clear that so long as C is entailed by T, the confirmation of T by C may well be inferentially misleading. Only when some component P of T has been identified which is confirmed by T's consequence C even though P does *not* entail C—in fact, only when neither C nor P entails an uncertain component of the other[41]—do we have evidence that verification of C gives any epistemically *significant* support to T. That verification of C is, in fact, often felt to confirm some such P in cases where C and P are ingredients of a tightly knit theory is undeniable, but the hypothetico-deductive account of nondemonstrative inference gives no clue whatsoever about the principles governing this flow of credibility.

The inductivist approach The juncture at which we have now arrived is this: Two scientific propositions P and C are often so related that verification of C increases the credibility of P even when C is *not* a logical consequence of P or conversely; and the foremost methatheoretical problem of scientific inference is, for now, simply to *learn* what patterns of propositional relationship effect this sort of credibility coupling. It is altogether possible—perhaps even necessary—

---

[39] If credibility can be represented by a probability measure, so that $\Pr(T)$, $\Pr(R)$ and $\Pr(C)$ are the probabilities (credibilities) of theory T and its components R and C, respectively, prior to verification of C, while $\Pr(T \mid C)$ and $\Pr(R \mid C)$ are the respective probabilities (credibilities) of T and R upon verification of C, we have $\Pr(T) = \Pr(C \cdot R) = \Pr(C) \times \Pr(R \mid C)$ while $\Pr(T \mid C) = \Pr(C \cdot R \mid C) = \Pr(R \mid C)$. Hence verification of C increases the credibility of T from $\Pr(T)$ to $\Pr(T \mid C)$ merely by replacing $\Pr(C)$ with unity.

[40] Since $\Pr(R) = 1 - \Pr(\sim R) = 1 - \Pr(C \cdot \sim T) = 1 - \Pr(C) \times \Pr(\sim T \mid C)$ while $\Pr(R \mid C) = \Pr(R \cdot C \mid C) = \Pr(T \mid C) = 1 - \Pr(\sim T \mid C)$, we have $\Pr(R \mid C) - \Pr(R) = -\Pr(\sim T \mid C)[1 - \Pr(C)] = -\Pr(\sim C) \times \Pr(\sim T \mid C) \leq 0$. Hence the probability (credibility) of R, given C, is less than the prior probability (credibility) of C unless either C conclusively verifies T or C's prior probability (credibility) is zero, in which case $\Pr(R \mid C) = \Pr(R)$.

[41] If both C and P have a common implicate H, each can be written as the conjunction of H with a residual. Hence verification of C includes verification of H, and if $\Pr(H) < 1$, confirmation of F through verification of C may be no more than a trivial result of verifying P's component H.

that such a connection between P and C involves a mediating theory T of which both P and C are consequences. (That is, while the confirmation of P by C is a *sign* that C confirms T nontrivially, what is given significant confirmation most immediately by C is in all likelihood something in T which implies C and from which P also follows.) However, our preceding argument has shown that for this to occur the relations among C, T and P must be more specific than *merely* the joint entailment of C and P by T. The problem would be obviated if there existed a satisfactory theory of propositional probability (or credibility) which supplied a computable unconditional probability (credibility) number for every proposition constructable in the science's language. We could then compute $\Pr(P \mid C)(= \Pr(P \cdot C)/\Pr(C))$ and $\Pr(P)$ and know not merely whether or not C confirms P but also, more importantly, exactly how much credence to invest in P upon verification of C. Despite the massive attention which probability theory has received within the past century by statisticians and philosophers, however, the only developments which have come within shouting distance of a practical theory of propositional probability are essentially confined to inferences about probability parameters (and whatever can in turn be inferred from these) based on sample frequency distribution; while to my knowledge no one has even begun to explore probability measures over theories which postulate non-observational entities. Over much—perhaps all—of the domain of non-demonstrative inference it seems less likely that propositional probability theory will furnish the grounds for drawing inferences from datum-beliefs as that will be a purification and unfolding of certain primitive induction patterns which are either themselves presupposed in the theory's axiomatic basis or are taken as criteria against which the theory's adequacy is assessed. That is, if verification of C confirms P, the explanation for this is not likely to be found in a more fundamental fact that $\Pr(P \mid C) > \Pr(P)$; instead the latter is most probably *due to* the former while the reason *why* C confirms P must be dug out of the structural relations between P and C. If so, the inductivist thesis has won the day.

The essential tenet of an inductivist interpretation of scientific inference is that there are, in fact, descriptive/normative formal patterns by which belief flows from one proposition (or set of propositions) to others not logically entailed by the former. If such patterns exist, they may be expected to guide datum-beliefs toward conclusions which complete the gestalt; hence inductivists are disposed to admit of a "logic of discovery" as well as of a "logic of confirmation." But the inductivist thesis does not require that all inductive conclusions be called to mind by the evidence which supports them. It insists only that no matter what causes the conclusion to be first thought of, its plausibility follows from recognition (perhaps intuitively, without explicit awareness) of its inductive relation to the evidence. What differentiates this from hypothetico-deductivism is that (a) inductivists, unlike hypothetico-deductive partisans, are disposed to look favourably upon the possible existence of a logic of discovery; and (b) inductivists are prepared to find that the relation between an inductively justified conclusion and its evidence depends upon logical structure a great deal more specific than mere entailment.

The state of contemporary theory on inductive logic is far too involved for me to say much about it here beyond that we still have a long, long way to go.[42] But it is perhaps not altogether misleading to summarize this work as consisting almost exclusively of attempted justification, reconstruction, and enrichment of the primitive inference schema of *statistical induction*. In brief, statistical induction is a projection of the pattern of things already known upon our expectations about things yet unencountered. Thus if $p\%$ of observed αs have been βs, we infer by statistical induction that probably about $p\%$ αs still forthcoming will also be βs. So characterized, this inference form is not completely general—in ways which are still distressingly obscure its intuitive applicability depends upon additional features of the projected pattern and the nature of the entities involved, so that artificial examples can easily be constructed in which the formal induction schema yields painfully counterintuitive conclusions (see Goodman, 1955). Even so, until very recently this was the only inductive argument form known to abstract metascience; in fact, throughout much of philosophy and logic the term "induction" is synonymous with "statistical induction." Unfortunately, statistical induction appears hopelessly inadequate to confer credibility upon hypotheses more complex than mere statistical generalizations,[43] which is one reason why the inductivist position has seemed simple-minded to many philosophers.[44] I have recently been able to show, however, that at least some inferences to unobserved (theoretical$_{\mathrm{B1}}$) entities are governed by a standard induction form in which an observed property of a class as a whole is transformed into an inferred attribute ascribed to each member of the class (Rozeboom, 1961)—though again, as in statistical induction, the intuitive acceptability of the argument requires special conditions, whose nature is unclear. That this is the only route by which we come to believe in underlying entities is most unlikely (in fact, see Rozeboom (1966b, p. 208ff.)); what is metatheoretically fundamental is that argument forms by which propositions incorporating theoretical$_{\mathrm{B1}}$ concepts originate and command observational support—i.e. patterns of *ontological induction*—do, in fact, exist. Hence there is

[42]For a valuable review of recent developments, see Kyberg (1964).

[43]The hypothetico-deductive argument can be construed as a special case of statistical induction if we reason that insomuch as all the consequences of theory T tested so far have been verified, all its remaining consequences are probably also sound. In fact, I suspect that this way of looking at the matter does occasionally have force in real-life theory assessments. However, statistical induction from the truth of tested consequences of T to the truth of yet untested consequences of T is marginal at best with respect to the intuitive restrictions (notably, something akin to "randomness" in the selection of observed cases) requisite for statistical induction to feel convincing.

[44]See, e.g. Wisdom (1952), and Bunge (1963).

no good reason to discount the possibility that all varieties of plausible inference, including justification for some of the most recondite and observationally remote components of scientific theories, instantiate one or another of certain determinate induction forms.

Degrees of credibility within theories

From this absurdly inadequate survey of the grounds for theory credibility, two implications nonetheless emerge which have major significance for practical theorizing in psychology. The first is that with the partial exception of inferences about statistical parameters, there still exist no accredited standards of non-demonstrative inference to which we can turn for normative assistance when judging the plausibility of a particular psychological theory. Not only are we strictly on our own in this respect, with no sagacious metatheoretical father-figure standing behind us to beam approval for our right thinking and gently correct our inferential errors, we have a double responsibility for special care in our evaluations of theories both to avoid inappropriate belief on the theory's own account *and* to provide instances of inductive arguments which have passed the test of detailed appraisal by professionally polished sensitivities in order that metatheory may eventually abstract from these the patterns which count as sound inference. Secondly, we have seen that a theory cannot properly be accepted or rejected holistically, but that the available evidence gives each of its conjunctive components its own separate degree of credibility. In particular, a detachable component of the theory which has no implications for any of the data upon which the theory rests is unlikely under natural circumstances to draw much support from these data, though neither can this possibility be altogether excluded. This is the situation I had in mind when I contrasted *warranted* theory with *speculative* theory, above, though the distinction is best conceived as a continuum rather than a dichotomy and applies more usefully to the various components within a given theory rather than to the theory as a whole. To state the point once more, if theory T is analysed as a conjunction of premises $P_1, \ldots, P_n$—and it is important to realize that there are many alternative ways to perform such a logical partition, some of which may have little resemblance to the conjunctive organization in which the theory is first conceived—then even if all the evidence, D, available so far is logically consistent with T and perhaps even includes verification of certain initially implausible consequences of T, the credibility given by D to T's component P_i may range anywhere from near-certainty (a highly *warranted* inference from D) through indifference to D (i.e., $\Pr(P_i \mid D) \approx \Pr(P_i)$, in which case, unless P_i's prior probability is high, P_i is still *speculative*) down to the possibility that P_i is highly unreasonable in light of D.

Now, a serious danger from theories whose components are not confirmationally homogeneous with respect to present or immediately portending evidence is that unless each component's contribution to the theory's datum-entailments is made explicit, we shall inevitably find our theoretical beliefs grossly misaligned with what our own good judgment would dictate were we but able to apply it. In particular, hypothetico-deductive holism encourages us both to accept untested, irrelevant or implausible components of a theory which has successfully accounted for the major research findings in its area of application so far, especially if the theory has brought off some unexpected predictions, and to scorn theories containing some conspicuously untenable or distasteful ingredient which in fact is no more than a superfluous embellishment on premises which service the theory's actual work load.[45] I contend that most psychological theories *do*, in fact, suffer badly from this intellectual malady, and that those portions of theories which carry the colour, excitement and challenge of a particular psychological perspective seldom have much overlap with those portions which are sustained by evidential support.

If my argument for the credibility partitioning of theories is accepted, what then? Well for one, it goes a long way toward resolving the sometimes bitter metatheoretical polarity which now exists over the role of theory in psychological science. It is no secret that many psychologists who consider the goal of science to be the attainment of *knowledge* deplore theories about the underlying sources of overt behaviour as tantamount to mysticism and fantasy, or disown their own theoretical contributions by characterizing them as "models" whose only function is to describe and perhaps occasionally predict empirical regularities. But the *credibility* contrast between warranted and speculative theory components must not be confused with the observational/theoretical contrast in *perceptual accessibility*. A presumptive empirical generalization can exceed its evidential basis just as wildly as can a speculation about unobservables, whereas the truly relevant components in a theory about the underlying sources of some empirical phenomenon may have as strong an inductive warrant as any statistical generalization at the datum level. On the other hand, insomuch as inquiry which does not seek to replace speculation and surmise with a harder intellectual coinage is unworthy of scientific respect, no theory component is exempt from the harsh credibility tests at which professional scientists, including research psychologists, have become adept, and until a theory has been stripped of its fat and flourishes down to an austere core which is *demanded* (i.e., strongly confirmed) by extant data it well deserves the second-class cognitive status which hard-nosed empiricists are disposed to grant it.[46] This is not to contend that speculative theorizing has no place in science. Quite the opposite—imagination, hunches, and inspired guessing remain a highway to the novel empirical discoveries without which scientific knowledge would remain static.

[45] This point has been nicely emphasized by Deutsch (1960, Ch. 1).

[46] Thus one can wholeheartedly accept the spirit of Turner's (1961) distinction between "Type I" (scientifically respectable) and "Type II" (scientifically disreputable) conjectures while protesting his co-ordination of it with the observational/theoretical distinction.

But it still remains essential to distinguish what is, in fact, well confirmed from what *might* possibly be true.

In addition to urging metatheoretical conciliation over the scientific respectability of theories, recognition that a theory is usually a very mixed bag of credibilities calls for an essentially new approach to practical theory development. So far as the theory's initial creation is concerned, anything goes—though statistical and ontological inductions from known data will undoubtedly remain a primary source of inspiration, preoccupation with plausibility in the discovery phase of theory building will only inhibit imagination. But once a set of premises has been found which successfully accommodate the available evidence, it is a waste of time and intellectual energy to take a stance on the global acceptability of these premises, or to set about testing the theory-as-a-whole in traditionally indiscriminate hypothetico-deductive fashion. The efficient procedure is to do a metatheoretical dissection of the theory in order to see what specific function is served by each of its ingredients, both for explaining extant data and for prediction of other phenomena relevant to research interests in this area. I know of no analytic routine which will parse a theory along lines having the greatest epistemic perspicuity, but the separable components disclosed in the theory by skillful factoring should fall roughly into four main categories: (1) Premises which, through statistical or ontological induction, are highly confirmed by present evidence. (2) Premises which entail and in turn are inductively implied by certain tentative empirical regularities whose confirmation by the extant data is still spotty. (3) Premises which have no relevance for the data obtained so far but do have significant implications for (and hence confirmational ties to) potentially demonstrable phenomena lying within this research area's range of concern. (4) Premises which are relevant only to possible phenomena lying outside of this research area's scope. For example, suppose that a theory (or "model") under consideration as a possible interpretation of certain learning phenomena presupposes that all subjects in the population studied have the same value of a certain theoretical parameter τ. If it turns out that only the population mean on τ makes a discernible difference for what the theory implies should happen when background conditions or procedures are standardized as in the experiments conducted so far, but that the population variance on τ is importantly related to how certain alterations in these background conditions should affect the phenomena under study, then the premise that $\mathrm{Var}(\tau) = 0$ is a Class 3 component of this theory. On the other hand, if the theory explains these phenomena in terms of a hypothesized physiological mechanism, the theory's implications for overt behaviour would be unaffected if it were cut back to postulation merely that *something* of unspecified nature is functionally related to the data variables in the manner assumed by the physiological hypothesis. Consequently, the theory's further premise that the underlying mechanism which manifests these functional properties also has such-and-such a physiological constitution has import only for certain types of physiological observation which are of no concern for psychology as such (though of course it is perfectly legitimate for psychologists to have physiological curiosities as well) and is hence a Class 4 component of the theory so long as the latter's intent is only to account for psychological phenomena.

While the various categories of theory components obviously shade off into one another, each has its own distinctive cognitive and methodological status. Classes 1 and 2 compose *warranted* theory, with Class 1 propositions constituting what may be thought of as "scientific knowledge" while Class 2 hypotheses are reasonable but still tentative conclusions which require further substantiation before they can aspire to knowledge stature. Theory components in Classes 3 and 4, on the other hand, are fatty *speculation* deposits within the theory's cognitive muscle. No matter how amusing, provocative, or comforting these may seem in the private meditations of individual scientists, they lay no serious claims to credibility and thus have no place in the science's public pronouncements. However, Class 3 speculations differ importantly from those in Class 4 in that the former generally direct attention to previously unsuspected dimensions of the phenomenon under study (even though research on this new aspect of the problem will more likely than not disprove the conjecture which led to its investigation), whereas Class 4 speculations are a froth of irrelevancies which give the theory spurious bulk. To be sure, as the science broadens the scope of its interests or as new fields of study spring up in interdisciplinary lacunae, a Class 4 hypothesis may receive Class 3 reassignment and acquire an outside chance of eventually graduating to *warranted* status. But it is intellectually fraudulent for a theory to include detachable premises which make a difference only for some extraneous area in which the theory's proponents undertake no serious research responsibilities.

The practical desirability of identifying the implicative/confirmational status of a theory's assorted components should be evident. Fundamentally, it is simply a matter of keeping possibilities in their proper credibility perspective. Speculative fantasies should not be allowed to usurp assent by riding piggyback on well-confirmed inductions, nor should our recognition of the latter be in turn undermined through implausibility-by-association. In particular, it is unprofessional to bicker over the holistic merits of a given theory when its advocates are justifiably impressed with its Class 1 and Class 2 contents while its critics justifiably distrust its Class 3 assumptions and scorn its Class 4 pretentions. An important corollary for theory-guided research is that only if the theory in question has been analysed in the way here envisioned can we have much assurance that our experimental program to test and correct the theory yields data which are relevant to the points ostensibly at issue—and if not, what they *are* relevant to. Little is to be learned, for example, from continuing to verify consequences of the theory which follow from premises which, upon analysis, can be seen to have already been highly confirmed (though of course there is always, a fair chance that soft spots may appear even in Class 1 conclusions); while if the heat of theoretical controversy centres around the Class 4 nature of underlying mechanisms, verification of consequences derived from Class 3 hypotheses fused with the former in the theory's original formulation will not help to resolve the dispute even though these findings may well break open vital new dimensions of warranted theory in this area. Similarly, a perceptive articulation of the inferential relations between a theory's variegated premises and the observations they subsume is essential to effective modification of the theory with accumulating evidence. Contrary to hypothetico-deductive holism, which sees a theory's evolution as a sequence of *de novo* inspirations in which every revision arises from the shambles of its disconfirmed predecessors by a new act of creative discovery, each separate slug of empirical evidence has its own specific target within the theory where it strikes with confirmational or disconfirmational impact, and the growth of scientific knowledge is a continual accretion of local additions and corrections, any one of which leaves the warranted components of theory in this area largely unaltered.[47] It would be most unfortunate if, say, experimental disproof of some of a theory's Class 3 conjectures were to be construed as evidence against Class 2 premises from which the former can be detached, especially if this were to terminate research designed to firm up the inductive basis for the latter. More generally, it is important to know whether a disconfirmatory research outcome conflicts with the theory's Class 1 or Class 2 components and hence violates the pattern of events that seemed to be emerging from previous data, or whether what has been discredited is merely a speculation which was never much more than a cognitively arbitrary if aesthetically appealing guess in the first place.

Of course, the most seemly research attitude is one in which we are aware of the alternative theoretical possibilities hovering behind the known phenomena in the area under investigation, and appreciate their various implications for observations yet unmade, but feel no special partiality toward any particular one of the competing prospects. This has always been the empiricist research ideal, but as attested in our recent history by the emphasis placed upon hypothesis testing in graduate research instruction, it is downgraded[48] by the hypothetico-deductive outlook and often seems to conflict with our felt need for understanding which goes

[47]While Kuhn (1962) has argued persuasively for the *revolutionary* character of major advances in scientific theory, I am prepared to argue that genuine upheavals of foundations seldom if ever occur. My contention is that the body of accumulated data inductively determines a logical structure which, within limits, must be embedded within any theory adequate to subsume these data, and that Kuhn-type "revolutions" are for the most part shifts in the Class 3 and Class 4 speculations with which this structure is fleshed out though such a shift may well require a psychologically revolutionary conceptual overhaul while the pressure to revolution may come from data which undermine the Class 3 components of the old theory even as they support Class 2 components in its successor.

[48]For a remarkable exhibition of this attitude, see Medawar (1964). Fortunately, the hypothesis-testing bias in psychological research now appears to be on the wane.

deeper than mere surface generalizations. The contemporary surge of interest in "models," with its repudiation of the personal involvement that "theories" seem to demand, is a commendable move back toward old fashioned open-minded neutrality, but it, too, shirks its intellectual responsibilities by failing to distinguish between what in the model is really true and what is just pretend. Given an appropriate implicative/confirmational factoring of a theory's propositional content, however, the proper balance between commitment and impartiality should emerge more or less automatically upon realization that the speculative components of theory can be manipulated—or ignored—with as much zest, playfulness or disdain as suits one's mood without insult to one's dedicated conviction, if it comes to that, in those aspects of the theory which have genuine evidential support. Class 3 conjectures are then no longer bastions of emotional allegiance to be attacked or defended by all-or-none hypothesis-testing experiments, but become expendable roadmaps to parametric studies that enrich, rather than convulse, the explanatory framework within which our research is conceived.

Coda

What should a psychological theory be? It should be *analysed*—exactingly, sensitively, and exhaustively.

References

Ayer, A. J. (1936). *Language, truth and logic.* London: Gollancz.

Boring, E. G. (1957). When is human behavior predetermined? *Science Monthly*, *84*, 189–196.

Bunge, E. G. (1963). *The myth of simplicity.* Englewood Cliffs, N.J.: Prentice-Hall.

Campbell, N. R. (1920). *Physics: The elements.* Cambridge: Cambridge University Press.

Carnap, R. (1956). The methodological character of theoretical concepts. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science* (Vol. 1). Minneapolis: University of Minnesota Press.

Carnap, R. (1961). On the use of hilbert's ϵ-operator in scientific theories. In A. Robinson (Ed.), *Essays in the foundations of mathematics.* Jerusalem: Manes Press.

Chapanis, A. (1961). Men, machines, and models. *American Psychologist*, *16*, 113–131.

Deutsch, J. A. (1960). *The structural basis of behavior.* Chicago: Chicago University Press.

Feigl, H. (1945). Operationism and scientific method. *Psychological Review*, *52*, 250–9.

Feigl, H. (1956). Some major issues and developments in the philosophy of science and of logical empiricism. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science, Vol. 1.* Minneapolis: University of Minnesota Press.

Feigl, H. (1958). The 'mental' and the 'physical'. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science, Vol. 2.* Minneapolis: University of Minnesota Press.

Feyerabend, P. K. (1963). How to be a good empiricist—a plea for tolerance in matters epistemological. In B. Baumarin (Ed.), *Philosophy of science: The Delaware seminar.* New York: Wiley.

Hanson, N. R. (1958). *Patterns of discovery.* Cambridge: Cambridge University Press.

Kaplan, A. (1964). *The conduct of inquiry.* San Francisco: Chandler.

Koch, S. (1954). Clark L. Hull. In W. K. Estes (Ed.), *Modern learning theory.* New York: Appleton-Century.

Koch, S. (1959). Epilog. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3). New York: McGraw-Hill.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: Chicago University Press.

Kyberg, H. E. (1964). Recent work in inductive logic. *American Philosophical Quarterly*, *1*, 249–97.

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*, 95–107.

Mandler, G. (1963). Comments on *P*rofessor *J*enkins' paper. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill.

Maxwell, G. (1962). The ontological status of theoretical entities. In H. Feigl & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science, Vol. 3.* Minneapolis: University of Minnesota Press.

Medawar, P. B. (1964). *Is the scientific paper fraudulent? Saturday Review*, August 1, 1964.

Miller, N. E. (1959). Liberalization of basic r-r concepts: extensions to conflict behavior, motivation and social learning. In S. Koch (Ed.), *Psychology: a study of a science* (Vol. 2). New York: McGraw-Hill.

Nagel, E. (1961). *The structure of science.* New York: New York: Harcourt, Brace & World.

Pavlov, I. P. (1928). *Lectures on conditioned reflexes.* New York: Liveright.

Rozeboom, W. W. (1958). "What is learned?"—an empirical enigma. *Psychological Review*, *65*, 22-33.

Rozeboom, W. W. (1960). Do stimuli elicit behavior?—a study in the logical foundations of behavioristics. *Philosophy of science*, *27*, 159–170.

Rozeboom, W. W. (1961). Ontological induction and the logical typology of scientific variables. *Philosophy of Science*, *28*, 337-377.

Rozeboom, W. W. (1962). The factual content of theoretical concepts. In H. Feigl & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (Vol. 3). Minneapolis: University of Minnesota Press.

Rozeboom, W. W. (1964). Of selection operators and semanticists. *Philosophy of Science*, *31*, 282–285.

Rozeboom, W. W. (1965). The concept of memory. *Psychological Record*, *15*, 329–368.

Rozeboom, W. W. (1966a). *Foundations of the theory of prediction.* Homewood, Illinois: The Dorsey Press.

Rozeboom, W. W. (1966b). Scaling theory and the nature of measurement. *Synthese*, *16*, 170–233.

Seward, J. P. (1955). The constancy of the I-V: A critique of intervening variables. *Psychological Review*, *62*, 155–68.

Toulmin, S. (1961). *Foresight and understanding.* London: Hutchinson.

Wisdom, J. W. (1952). *Foundations of inference in natural science.* London: Methuen.

Workman, R. W. (1964). What makes and explanation? *Philosophy of Science*, *31*, 241–254.