Published in Feigl, H. and Maxwell, G. (eds) (1961) *Current issues in the philosophy of science.* New York: Holt, Rinehart, & Winston. (pp. 473–484)

Formal Analysis and the Language of Behavior Theory

Psychology today has a curiously schizoid attitude toward formal methodology and philosophy of science. On the one hand, we find a great deal of enthusiastic, if unskilled, discussion *about* methodology, particularly about general issues having little immediate implication for psychological research. On the other hand, any attempt to *do* methodology in the course of an actual research problem is likely to meet only indifference, incomprehension, and at times open hostility. This rejection of formal methodology at the working level of psychology is only too clear in the current difficulty in finding publication outlets in the psychological literature for discussions of nontraditional methodology arising from specifically psychological problems, and in the recent unanimous agreement by a distinguished panel of research psychologists that "philosophy of science has little or nothing to do with how research gets done in psychology" (Garner, Hunt, & Taylor, 1959).[1]

But this attitude is really quite puzzling. For scientific methodology, after all, is only the study of scientific method, and above all, it is *method* that makes science *science*, in contrast to such sources of belief as folklore, superstition, and the like. Methodological analysis of a research problem is little more than being aware of what one is doing and systematically looking for better ways to do it. In particular, *formal* methodology is merely the study of the scientist's verbal, or conceptual, techniques; and any researcher who feels that the unexamined language habits he picked up at his mother's knee are adequate to see him through the most complex scientific investigation is living at about the same level of reality as the high school pitching star who is sure he would be a valuable asset in the current major league pennant race. Moreover, where is the necessary methodology going to get done if not in close association with actual scientific problems? If the researcher sometimes feels that the dicta of philosophers of science are frequently oversimplified, excessively restrictive, and of little relevance to his own work, he has only himself to blame for making no attempt to clarify just what his conceptual procedures and objectives actually are. Hence the contention that behavioral psychology stands to profit greatly from careful scrutiny of its language techniques is really little more than a truism.

[1]Professor Hunt has informed me that the quoted statement was in no way intended to be prescriptive, but merely a description of the state of affairs actually found in contemporary psychological research.

Now, the formal methodology of science is not all of a piece. In particular, there are at least three problem-areas which, though interrelated, warrant separate recognition. I refer to the problem of evidence, the problem of concepts, and the problem of formal structure. Concern with the first, the problem of evidence, is as old as science itself—in fact, science as we now understand it originated when men began to insist that speculations about natural phenomena be supported by empirical evidence. The role of data as the foundation of the scientific enterprise is widely appreciated today, so I shall say no more about the problem of evidence here, though it is perhaps worth noting that only within the last two hundred years has a systematic theory of evidence begun to develop.

The problem of concepts, on the other hand, is of much more recent recognition. To be sure, it has been known for a long time that some concepts turn out to be more fruitful than others, and the experimentalist has always as a matter of course favored the precision of suitably defined technical terms over the vagueness of ordinary language, but only in quite recent times has the idea emerged that there are definite criteria to which a concept must conform if it is to be of use to science. Just what these criteria may be is still very much a controversial matter. The Spartan standards of militant operationism and positivism in the thirties are now yielding to more tolerant views of scientific meaningfulness. Still basic to the theory of concepts, however, is the principle that a term is scientifically acceptable only to the extent that its usage can be made precise. And while entirely too few psychologists subject their working vocabulary to the kind of intensive scrutiny and purification necessary for a science to become exact, still I think that except for a few die-hard intuitionists, most psychologists today would concede that conceptual analysis is a valid and desirable procedure, even if they themselves do not care to engage in it.

My primary concern today, however, is with a third kind of methodological problem which as yet has received little or no recognition in psychology—namely, that of formal structure. By the "theory of formal structure," I mean the systematic study of the logical forms of scientific assertions and, derivatively, the formal properties of the symbolic elements employed. Examples of items of interest under the latter heading are the representation of forces in physics, and colors in sensory psychology, by ordered triads of numbers, and the quantum-mechanical representation of the momentary velocity of a particle by a function, rather than by a scalar. The significance of such devices, however, lies in the manner in which they are embedded in the propositional network within which they occur, and it is this more basic question about the logical form of a set of assertions on which I now wish to focus attention.

In brief, to analyze the logical form of a system of statements is to identify the constituent concepts, note their logical types, and determine the formal schema by which they have been compounded. This is similar to a grammatical parsing

of the system except for the crucial difference that the grammar of a living language, English in particular, conceals many a vital conceptual element and logical relationship behind a facile stream of idiom, metaphor, and stylistic variation. A logical analysis of the sort described is practically a formalization of the system and, in fact, the primary objective of formalizing a set of propositions is not, as might be supposed, to make possible rigorous deductions, though this is always a useful consequence, but to bring out the logical structure of the set.

It is my intense conviction that it is just as important for the scientist to be explicit about the formal structure of his assertions as it is for him to be tough minded about his concepts and his evidence. Unfortunately, the typical research psychologist, who regards language technique more as an art form than a research instrument if he thinks about it at all, is normally about as interested in this sort of analysis of his working hypotheses as a professional quarterback would be in a ballet dancer's assessment of his broken-field running. But this attitude is based on a misunderstanding. Logical structure has nothing to do with linguistic style. The whole point of formal analysis is that assertions made in informal English, even when fortified with technical terms and special notations, simply are not *clear*. Over and above any vagueness residing in its individual terms, unformalized speech teems with condensations, abbreviations, ellipses, suppressed premises, etc., which serve to conceal a substantial portion of what is actually being said. A goodly part of the analytic philosopher's livelihood is earned "unpacking" the hidden, and hence frequently unsupported, premises that can usually be found tucked away within some grammatical subtlety or meaning shade of even the most innocuous appearing sentence; and while sentences for which such analysis has been carried through have been for the most part those which crop up in typically philosophical disputes, there is no reason to think that a science which relies on this same linguistic machinery fares much better. Nor is mathematization of the science likely to increase its logical clarity to any appreciable degree—quite the contrary, a chief virtue of mathematical formulae is their extraordinarily compact condensation of ideas which would be of horrendous complexity if written out in logical fullness. In any case, the philosophically untrained scientist is seldom aware of the full content of the statements he subscribes to, whether these be finely spun theories or supposedly simple generalizations from known data. The research scientist, in particular, is jeopardized by the hidden commitments of his language; for what he believes to be little more than a summary of his observations may in fact sneak in further assumptions which not only are unwarranted by the data at hand, but may be in conflict with other known data or established belief. If time permitted, I would show how one of the great disputes of modern learning theory—namely, What does the organism learn during conditioning?—has been a controversy of precisely this kind.[2] When an apparent clash of data

[2]See Rozeboom (1958).

arises in this way, the solution is not to rush back to the laboratory, but to look for methodological therapy. When impacted ideas become infected, extraction is necessary, and the instrument for this is formal analysis.

To illustrate the concept of formal structure and the concealments of ordinary discourse, let us look at a couple of transparent examples which in themselves are unlikely to generate any puzzles, but which exhibit some of the ways in which serious difficulties do, in fact, arise in more involved contexts.

Consider the assertion,

The human ovum contains 23 chromosomes.

Statements of this form are prevalent in textbook introductions to genetics, and it is unlikely that any geneticist would take offense at it. Yet taken literally, this statement is grievously in error. Grammatically, it is in singular form, ascribing a certain property to a unique object answering to the definite description, "the human ovum." Hence as it stands, this assertion presupposes either that there is just one human ovum, or that there is somewhere a single prototypical "true" human ovum, a platonic ideal of which the billions of existent human ova are merely imitations. But, of course, a person who claims that the human ovum contains 23 chromosomes would never accept this literal interpretation of his statement. Roughly, what he means is that every human ovum contains 23 chromosomes. And yet, put baldly, the claim,

All human ova contain 23 chromosomes

does not sit comfortably either, for meiosis does go wrong occasionally, and we can feel sure that an occasional ovum arises with more or less than 23 chromosomes. In any case, the proper logical form of the proposition which "The human ovum contains 23 chromosomes" is intended to convey is that of an assertion about the frequency of a characteristic within a class, which is something quite different from what the uttered sentence actually *does* say. Moreover, a person who uses this singular form to express his beliefs about chromosome number is thereby vulnerable to a particularly insidious form of intellectual blindness. He is not so naive as to be led into assuming that there is only one genuine human ovum—he will readily admit that "The human ovum contains 23 chromosomes" is shorthand for a statement about chromosome frequency in the class of human ova. But this singular-form shorthand is unable to convey any information about the incidence of atypical cases, or even to suggest the possibility thereof, and hence use of this formally improper abbreviation carries with it the implicit assumption that *all* human ova have 23 chromosomes. That is, so long as a person thinks about genetics in terms of the singular sentence-schema,

The ovum of species s contains x chromosomes,

he will be blind to the problem of the *distribution* of chromosome number within the ova of a given species, and to the challenging research possibilities attendant upon the abnormal case. Conversely, if the logical form of the belief which lies implicit in the use of the singular schema is made explicit—namely,

All ova of species s contain x chromosomes,

attention is immediately drawn to the uncompromising logical operator "all," and it becomes obvious that a theory of genetics which does not allow for some variation in chromosome number within a species is an oversimplification, as shown in practice by its inability to account for certain genetic anomalies which have been observed empirically.[3]

For a second illustration of the problem of formal structure, consider the commonplace belief,

Adequate diet is necessary for good health.

This is probably true, and overlooking the horrible vagueness of "adequate diet" and "good health," few persons would have trouble understanding it. But suppose that certain poorly trained physicians were using this maxim to justify their practice of telling mothers to make a special point of eating well when their children are ill, or of paying no further attention to a patient's diet because he ate well last year. These inferences seem absurd-but why? On what grounds can we argue that the assertion, "Adequate diet is necessary for good health," does not warrant concern for the food habits of the mother of a sickly child, but *does* warrant concern for the diet of the child himself? The answer begins to emerge when we look at the literal meaning of this assertion. What it says is that a relationship of causal dependence holds between two abstract entities, adequate diet and good health. But this is nonsense, a "category mistake"; the grammar of concepts about causal relations admits their application only to states of affairs—that is, to situations, facts, or events. For example, while it is highly appropriate to ask whether the fact that Tommy broke out in red spots today is due to his having been exposed to chicken pox last week, it makes no literal sense to ask whether the attribute Breaking-out-in-red-spots is due to the attribute Being-exposed-to-chicken-pox. It is clear that since the formula, "Adequate diet is necessary for good health," does seem to be empirically meaningful, it must be elliptical for a more formally complex statement expressing a connection between certain facts or events involving adequate diet and certain others involving good health. Further reflection shows that what it is *meant* to convey is that for a person to be in good health during a certain period of time, it is necessary for that same person to have had an adequate

[3]Supermales and superfemales in *Drosophila*, for example.

diet during a slightly antecedent period. This explains why the physicians' advice cited above seemed so peculiar. It also forces us to recognize that we really are not very clear about the precise temporal relations involved. Quite typically, the ellipsis, "Adequate diet is necessary for good health," manages to suppress not only important elements of logical structure, but also awareness of corresponding uncertainties in our knowledge.

Let us now turn to a similar but much more damaging ellipsis in the language of behavioral psychology. The fundamental tenet of behavioristics is that given sufficient background data about an organism—that is, its conditioning history, body-maintenance state, species, etc.—the behavior of that organism can be predicted with a fair degree of success from suitable information about its immediate environment. In the jargon of behaviorese, this is put by saying that under suitable conditions, stimuli elicit or evoke responses, or, less forcefully, that stimuli predispose, set the occasion for, or "cue" certain behaviors. We say, for example, that food elicits salivation, that sudden loud noise may evoke a fear reaction, that under certain training antecedents the sound of a buzzer will predispose lever pressing, or that a red square is a cue for turning left. In reporting a conditioning experiment, it might be stated that the conditioned stimulus, the sound of a bell, was conditioned to a pupillary response through repeated pairings of the CS with the unconditioned stimulus, a bright light. Yet literally, all this is nonsense in precisely the same way that "Adequate diet is necessary for good health" is nonsense. For while the grammar of stimulus terms is not infrequently ambiguous, a little analysis shows that from the standpoint of logical type, stimuli are essentially attributes or thing-kinds (for example, red square, sound of buzzer, food, lever), though sometimes particulars (for example, the experimenter, a particular stimulus card, a particular lever, etc.) are also regarded as stimuli. Similarly, responses are momentary attributes of the organism (for example, salivating, turning left, pressing lever). But to speak of "eliciting," "evoking," "predisposing," etc., is to make reference to some sort of causal efficacy. Hence to say that a certain stimulus elicits or predisposes a certain response is to predicate of attributes or the like a relation that logically can hold only between facts or events. Thus in Pavlov's classic experiment, the abstract entity Salivation was not elicited by another abstract entity Meat-powder; rather, the blowing of meat powder into a certain dog's mouth at a certain moment caused that dog to salivate a moment later.

In short, just as the earlier maxim about diet and health must be expanded into a more logically complex statement concerning a lawful connection between events of certain kinds, so the standard formula of behavior theory—namely, that under conditions C, stimulus S elicits response R—must be interpreted as shorthand for a statement connecting certain events involving stimulus S with certain other events involving stimulus R. Now after a fashion, behaviorists are actually aware of this. A not infrequent alternative to the "Stimulus S elicits response R" formula

is "The *occurrence* of S elicits an *occurrence* of R"—thereby recognizing that it is the fact that the attribute or thing-kind S occurs in the presence of a certain organism at a certain moment which (in part) causes that organism to perform R a little later. The trouble is, this is the *only* reasonable expansion that can be given to the "S elicits R" ellipsis. The result is that just as use of the formula, "The ovum of species s has x chromosomes," implicitly assumes that all ova of a given species have the same number of chromosomes, thinking about elicitation phenomena in terms of the formula, "Stimulus S elicits response R," generates the implicit assumption that only environmental events described by the sentence-schema,

Stimulus S is present to organism o at time t,

are to be regarded as potential elicitors of behavior. The effect of this in actual practice has been that prediction of an organism's behavior in a certain environmental situation is seen by most behaviorists as a matter of determining what stimulus elements occur in that situation, and then predicting a more-or-less algebraic result of the response tendencies associated with these various stimuli for an organism of this species, past experience, etc. That is, use of the traditional S-R ellipsis has surreptitiously committed behaviorists to the assumption that those facts of the environment which elicit or predispose an organism's behavior can adequately be described for purposes of behavioral prediction by a simple *listing* of the attributes, thing-kinds, or other simple characteristics present in the environment; for example, that if the relevant environmental events are that a light is flashed, food is displayed, and a bell is rung, it suffices to represent the eliciting situation by the list of terms, "Light-flash," "Food," and "Sound of bell."

But except for special cases, this just will not do. It is certainly not generally the case that the causal import of an event can be determined from a list of its components. For example, the implications of Rover's biting the mailman yesterday are quite different from those of the mailman's biting Rover yesterday, even though the ingredients of these two states of affairs are the same-namely, Rover, the mailman, yesterday, and the relation of biting. The *way* in which the components are combined—that is, the *structure* of the event—is also a crucial aspect, and this can be preserved only by describing the situation by a *sentence* (or its equivalent), not by a mere list of some of the terms which occur in the sentence. Similarly, tabulation of the stimulus elements present in an organism's environment is in general inadequate to capture the behavioral significance of that situation, because it is not always the mere presence of those stimuli which elicit or predispose the behavior—the structure of the eliciting events may also be relevant.

This point may be illustrated with a problem in discrimination learning which historically has proved rather embarrassing to traditional S-R formulations. There is good reason to think that when organisms are given repeated choices between

various pairs of objects colored G_1 and G_2, respectively, where G_1 and G_2 are shades of grey such that G_2 is darker than G_1, and are rewarded for choosing the G_2-colored member of the pair, some acquire what may informally be described as a preference, not for G_2-colored objects, but for the darker of two grey objects, as shown (in part) by the organism's behavior when now given a choice between a G2-colored object and one of an even darker shade. Let an organism that has acquired such a response tendency be said to have habit H. Now, what technical analysis is to be given the behavioral situation in which an organism o with habit H is confronted at time t with a pair of objects colored G_i and G_j respectively, where G_i and G_j are arbitrarily selected shades of grey? In particular, what are the behaviorally relevant aspects of the eliciting situation, what is the resulting behavior, and especially, what is the generalization by which a specific response event is predicted from a specific environmental situation? No serious discussion of this problem is possible in the time here allotted, but perhaps an intimation can be given of the major themes. The traditional behavioral approach, steeped in the "Stimulus S elicits response R" ellipsis, sees the eliciting event as the occurrence of the colors G_i and G_j in the presence of o at time t, a situation which can just as effectively, albeit elliptically, be described by saying that the elicitors are two stimuli, G_i and G_j. Prediction of a choice between G_i and G_j by o at t then proceeds by way of a highly complicated generalization about the comparative strengths of approach tendencies to various shades of grey derived from a set of generalization gradients, the predictive consequences of which alter violently both with small variations in the details of the learning history and with certain rather *ad hoc* theoretical assumptions. Whether or not such an approach succeeds in making reasonably accurate predictions (and even if parameters can be chosen to give suitable results for stimulus situations sufficiently restricted in kind, the whole line of reasoning collapses if the context of analysis is broadened), the procedure very much resembles the Ptolemaic approximations to the planetary orbits by series of cycles and epicycles. On the other hand, once we stop trying to describe the eliciting situation as a mere set of stimuli, it is quite simple to express the color-choice behavior of an organism with habit H. The generalization is that if o has habit H at time t, then, for any grey objects x and y, if o is confronted with x and y at time t and x is darker than y, then o prefers x to y at time t.[4] Then, for a specific situation, if A and B are grey objects present to o at t such that A is darker than B, the primary eliciting event is the *fact* that A is darker than B—a situation which can only be described by an expression with the syntactical structure of a sentence—and from which it is elementary to conclude, via the generalization, that o will prefer A to B. Note, by the way, that it is not at all correct to drag in the "S elicits R" formula, as a number of psychologists

[4]Of course, the terms 'confronted' and 'prefers' need to be given an adequate behavioral definition.

have tried to do, by saying that *o* responds *to the relationship*—that is, that *o* has acquired a response to a stimulus element, "Being-darker-than." The entire structured fact that *A* is darker than *B* must be made use of in predicting the choice.

In closing, I would like to indicate, as best I can in a sentence or two, how clear recognition that the elicitors of behavior are *events*, not just the stimulus components of those events, makes sense out of, and brings behavior theory into formal harmony with, certain more "tender-minded" issues in psychology. I have particularly in mind the "holistic" approach and the problem of cognitive states.

The term "holism" has come to refer to a rather diffuse attitude toward natural phenomena according to which nature can adequately be discussed only as a totality; that is, that a whole is greater than the sum of its parts, and that any analytic fragmenting of nature necessarily misses the truly important aspects. In psychology, the "holistic" approach has been particularly associated with the Gestalt movement and with respect to behavior takes the form that what the organism responds to is not disconnected pieces of its environment, but the "situation as a whole." Until recently, I regarded sentiments of this sort to be obscurantistic gibberish, and I still wonder whether partisans of holistic doctrine have more than a foggy notion of what they are trying to say. But it now occurs to me that if what the holist is protesting against is the view that the behavioral significance of a situation can adequately be captured by a simple compiling, no matter how exhaustive, of the objects, attributes, relations, etc., involved in the situation, then to this extent his position is unassailable: It is not grammatically possible for a mere list of terms to describe a situation; the description must have the kind of structural integrity found in a *sentence*. In order to predict behavior successfully, we must be given *facts* about the environment, not just a set of stimuli. (Of course, the listing of stimuli present does serve to convey *some* factual information about the environment. Further, if the holist presses his luck by interpreting his "situation as a whole" to be the environment in its *entirety*, his stand becomes simply a denial of a science of behavior. We have no reason at all to doubt that a few well-chosen facts about the circumstances in which an organism finds itself, given enough information about the organism, will serve quite adequately for behavioral prediction.)

Finally, a word about behavioristics and cognitive states. The cognitive functions of higher (and lower?) organisms, including such matters as thinking, believing, perceiving, semantical processes, etc., have so far remained essentially opaque to behavioral insight. Part of the trouble, of course, is that these are internal processes of a kind that behavioral psychology once hoped to eschew. Yet for some time, now, behavior theory has employed theoretical constructs with internal reference, and which have, in fact, been proposed as the behavioral counterparts of the so-called "higher mental processes." Now, there are certain important observations

which need to be made about cognitions—that is, about perceptions and beliefs. The first is that an organism's perceptions and beliefs are *action enjoining*—an organism o does R_i, rather than R_j at time t because it perceives or believes P, rather than Q, at time t. Secondly, in some important sense, perceptions and beliefs *correspond*, rightly or wrongly, to aspects of the world outside the organism, and it is only by means of these internal representations that the environment is able to incite and direct voluntary behavior. It follows that there must be a formal isomorphism between our descriptions of cognitions and our descriptions of the action-enjoining aspects of the environment. Now as has already been discussed, traditional behavioristics has been seduced by the S-R ellipsis into talking as though the environmental arousers of behavior are stimuli, and sure enough, the concepts which this approach has felt it necessary to introduce in reference to processes mediating between stimulus input and response output, such as "response surrogate," "r_g," "mediation response," etc., have the same formal status as stimulus concepts—namely, that of simple terms. But ordinary discourse and mentalistic psychology make abundantly clear that beliefs and also, though less obviously, perceptions, are *propositional* in nature.[5] That is, the proper logical form for ascribing a perception or belief to someone is "Person o at time t perceives (believes) that ____," where the blank is filled with a *sentence*. Moreover, when one attempts to trace the behavioral import of perceptions and beliefs, this formal complexity is seen to play an essential role. Hence the traditional S-R gropings toward a theory of cognition are doomed at the outset by their inadequate formal structure. But this defect issues only from a linguistic artifact, not from any intrinsic inadequacy in the behavioral approach. Once we adjust the conceptual framework of behavioristics to make explicit recognition that facts, not stimuli, elicit behavior, it will follow almost automatically that the behavioral description of internal states which cognitively represent these elicitors will have the necessary formal complexity, and the chasm which now gapes so forbiddingly between mentalistic and behavioristic psychologies will have yielded to an effortless rapport.

References

Garner, R. W., Hunt, H. F., & Taylor, D. W. (1959). Education for research in psychology. *American Psychologist*, *14*, 167–179.

O'Neal, S. M. (1958). Basic issues in perceptual theory. *Psychological Review*, *65*, 348–361.

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

[5]The propositional nature of perception has been stressed by O'Neal (1958)