Meehl, P. E. (1970). Psychological determinism and human rationality: A psychologist's reactions to Professor Karl Popper's "Of clouds and clocks." In M. Radner & S. Winokur (Eds.), *Minnesota Studies in the philosophy of science: Vol. IV. Analyses of theories and methods of physics and psychology* (pp. 310-372). Minneapolis: University of Minnesota Press.

Psychological Determinism and Human Rationality: A Psychologist's Reactions to Professor Karl Popper's "Of Clouds and Clocks"

Paul E. Meehl

In the Second Arthur Holly Compton Memorial Lecture, engagingly titled "Of Clouds and Clocks," Sir Karl Popper (1966; all quotations of Popper's language are from this source) addresses himself to a long-familiar problem about psychological determinism, indicated by the lecture's subtitle, "An Approach to the Problem of Rationality and the Freedom of Man." The lecture treats of several interconnected themes, ontological, historical, and methodological. I want to emphasize that the present paper is in no sense an "attack" on the lecture as a whole, which abounds with the usual Popper stimulation and perspicuity, and from which I have learned much. Some of the interpretation (e.g., the indeterministic features of classical physics) is beyond my competence even to discuss, let alone criticize. What I consider herein, qua philosophically oriented psychologist, is only one specific thesis, to wit, *psychological determinism is incompatible with human rationality*. The core idea here, in spite of the new aspects illuminated by Popper, is an old one, no doubt familiar in one form or other to almost any undergraduate philosophy major. (I recall first hearing it, when a sophomore, forcefully presented and ingeniously defended by Professor Alburey Castell. I did not buy it then, and I find that, some thirty years having passed, I cannot buy it now. But I trust that my reasons are somewhat better today than they were in 1939, as I am now more cognizant of the genuine puzzles and paradoxes involved.)

It is important to be clear about three matters right off: First, the thesis is ontological, not epistemological, and I therefore bypass evidential questions, freely invoking "what Omniscient Jones knows," "what a Utopian physiologist would say," "what is actually going on," "what is true concerning the state of Nature." The thesis is a claim about incoherence in a deterministic ontology; it says that if all human thought and action were completely determined, then it *could not be* rational. That kind of question can of course be examined without reference to the evidential issues of how we could find out that determinism was false, or how we could ascertain that we are rational (?!).

Second, I do not attempt a defense of complete psychological determinism, partly because its truth or falsity would not bear on its consistency with rationality, but also because I am not myself a convinced determinist, and

AUTHOR'S NOTE: I am indebted to the Louis W. and Maud Hill Family Foundation and to the Carnegie Corporation of New York for support through summer appointments as professor in the Minnesota Center for Philosophy of Science.

consider the substantive issue in doubt on present evidence.

Third, being a psychologist I am naturally suspect and vulnerable to a kind of *ad hominem* complaint that "You of course defend determinism because of trade-union interests, thinking that your scientific and clinical jobs require an implicit faith in the ultimate strict orderliness of all psychological processes." For me, at least, this is not true. I anticipate that no development of the behavior sciences will eliminate their current stochastic features, and I am not aware of any research programs that would have to be abandoned as fruitless if an element of radical indeterminism were postulated. For example, it seems fair to say that the greatest degree of behavioral prediction and control achieved thus far by psychologists is found in the work of Skinner and his disciples.[1] Aside from the fact that these modest triumphs of "behavioral engineering" are quantitatively tighter when the subjects are pigeons pecking keys than when they are humans speaking words—let alone philosophers engaged in criticism—whether or not one sees the laws as deterministic depends upon the level of analysis. The main dependent variable studied is rate of responding, as represented by the slope of a cumulative response record. The conceptual and mathematical relationships between this "operational" variable and the underlying probability-of-responding—the relation between a finite relative frequency and a "propensity"—have never been precisely explicated by the operant behaviorists. Usually they can go along quite well with their work without a rigorous explication of it. But when radical determinism is under discussion, we need more than a mere showing that a response-curve slope is highly manipulable. Whether or not a rat, pigeon, or human emits a certain response during a small interval Δt, and whether the response has such-and-such narrowly specified topographic, durational, and intensive properties, are *not* under complete experimental control, but remain probabilistic only. Besides, the various kinds of human psychological activity differ in how "clocklike" versus "cloud-like" they are, and Popper could quite properly argue that any showing that there is quasi-clocklike orderliness in a well-conditioned eye-blink reflex is only faintly relevant to the question "How clocklike are political theorizing, mathematical invention, and philosophical criticism?"

With these disclaimers made about what I am not attempting to do, what I

---

[1] In recognizing the unblinkable fact that Skinner's epoch-making book *The Behavior of Organisms* (1938) gave rise to a technology of behavior control which has, to an unprejudiced mind, no real competitors, I do not commit myself as regards its long-term theoretical adequacy. As would be true for most of my philosophical readers, I have grave reservations about Skinner's account of language, in his *Verbal Behavior* (1957), vigorously attacked by Chomsky (1959; but see MacCorquodale, 1970). I also believe the Skinnerian group underestimates the importance of genetic factors—and resulting real taxonomic entities—in behavior disorder, and hence they underrate the importance of formal diagnosis, although there is nothing about the theoretical position that requires this attitude. Finally, as a psychotherapist I have reservations about the adequacy with which Freud's constructs are translated into behaviorese by Holland and Skinner in *The Analysis of Behavior* (1961), a programmed text which I recommend to readers approaching this subject matter for the first time. Another good introductory presentation can be found in Skinner's *Science and Human Behavior* (1951). See also, but requiring varying amounts of technical preparation, Ayllon & Azrin (1968); Catania (1968); Ferster & Skinner (1957); Honig (1966); Krasner & Ullmann (1965); Skinner (1961, 1968); Ullmann & Krasner (1965); Ulrich, Stachnik & Mabry (1966); and Verhave (1966).

shall attempt is to criticize Popper's view that *if* human thought and behavior were completely determined, *then* they could not be rational. If I understand him rightly, he believes that if strict determinism were true, we could not, in any genuine sense, give reasons, be influenced by reasons, engage in critical thought, etc., and that the validity or invalidity of arguments could not influence the course of human happenings. The line of his argument can best be seen from a few representative quotations. Popper writes:

> [Quoting Compton] "If . . . the atoms of our bodies follow physical laws as immutable as the motions of the planets, why try? What difference can it make how great the effort if our actions are already predetermined by mechanical laws . . .*?"*
>
> Compton describes here what I shall call '*the nightmare of the physical determinist*.' A deterministic physical clockwork mechanism is, above all, completely self-contained: in the perfect deterministic physical world there is simply no room for any outside intervention. Everything that happens in such a world is physically predetermined, including all our movements and therefore all our actions. Thus all our thoughts, feelings, and efforts can have no practical influence upon what happens in the physical world: they are, if not mere illusions, at best superfluous by-products ('epiphenomena') of physical events. [pp. 7-8]

> I believe that the only form of the problem of determinism which is worth discussing seriously is exactly that problem which worried Compton: the problem which arises from a physical theory which describes the world as a *physically complete* or a *physically closed* system. By a physically closed system I mean a set or system of physical entities, such as atoms or elementary particles or physical forces or fields of forces, which interact with each other—and *only* with each other—in accordance with definite laws of interaction that do not leave any room for interaction with, or interference by, anything outside that closed set or system of physical entities. It is this 'closure' of the system that creates the deterministic nightmare. [p. 8]

> For according to determinism, any theories—such as, say, determinism—are held because of a certain physical structure of the holder (perhaps of his brain). Accordingly we are deceiving ourselves (and are physically so determined as to deceive ourselves) whenever we believe that there are such things as arguments or reasons which make us accept determinism. Or in other words, physical determinism is a theory which, if it is true, is not arguable, since it must explain all our reactions, including what appear to us as beliefs based on arguments, as due to *purely physical conditions.* Purely physical conditions, including our physical environment, make us say or accept whatever we say or accept; and a well-trained physicist who docs not know any French, and who has never heard of determinism, would be able to predict what a French determinist would say in a French discussion on determinism; and of course also what his indeterminist opponent would say. But this means that if we believe that we have accepted a theory like determinism because we were swayed by the logical force of certain arguments, then we are deceiving ourselves, according to physical determinism; or more precisely, we are in a physical condition which determines us to deceive ourselves. [p. 11]

> For if we accept a theory of evolution (such as Darwin's) then even if we remain sceptical about the theory that life emerged from inorganic matter we can hardly deny that there must have been a time when abstract and non-physical entities, such as reasons and arguments and scientific knowledge, and abstract rules, such as rules for

> building railways or bulldozers or sputniks or, say, rules of grammar or of counterpoint, did not exist, or at any rate had no effect upon the physical universe. It is difficult to understand how the physical universe could produce abstract entities such as rules, and then could come under the influence of these rules, so that these rules in their turn could exert very palpable effects upon the physical universe.
>
> There is, however, at least one perhaps somewhat evasive but at any rate easy way out of this difficulty. We can simply deny that these abstract entities exist and that they can influence the physical universe. And we can assert that what do exist are our brains, and that these are machines like computers; that the allegedly abstract rules are physical entities, exactly like the concrete physical punch-cards by which we 'program' our computers; and that the existence of anything non-physical is just 'an illusion,' perhaps, and at any rate unimportant, since everything would go on as it does even if there were no such illusions. [p. 12]

> For obviously what we want is to understand how such non-physical things as *purposes*, *deliberations*, *plans*, *decisions*, *theories*, *intentions*, and *values*, can play a part in bringing about physical changes in the physical world. [p. 15]

> Retaining Compton's own behaviorist terminology, Compton's problem may be described as the problem of the influence of the *universe of abstract meanings* upon human behavior (and thereby upon the physical universe). Here 'universe of meanings' is a shorthand term comprising such diverse things as promises, aims, and various kinds of rules, such as rules of grammar, or of polite behavior, or of logic, or of chess, or of counterpoint; also such things as scientific publications (and other publications); appeals to our sense of justice or generosity; or to our artistic appreciation; and so on, almost *ad infinitum*. [p. 16]

I believe these quotations suffice to give the essential argument, which purports to show that complete psychological determinism, arising on the basis of complete brain-process determinism ("mind is a function of brain"), renders genuine rationality and purposiveness impossible, and "giving of reasons" a spurious idea or an inefficacious irrelevancy. I turn now to my analysis and criticism of that contention.

There is, at least prima facie, a certain oddity about the position of those who wish to reject psychological determinism on the ground that it precludes human rationality, since part of their reason for insisting upon a "something else" which is not the mere workings of the cerebral machinery is the obvious fact that our conduct *is* causally influenced by "the giving of reasons." If I want to control your behavior with regard to a certain decision, it is true that I may proceed by various kinds of irrational appeals (e.g., after the manner of Hitler); but it is also true that I may proceed by giving you what I believe to be good reasons for your behaving in the way that I desire. In fact, if I know you fairly well and believe you to be a highly rational man, I may well operate on the assumption that the *most* effective way to control your behavior is to present you with good reasons. Thus we have a situation in which the idea of *control*, or the determination of one event (your action) by means of introducing another event (my giving you good reasons), with reliance upon a kind of regularity ("Jones is influenceable by good reasons," roughly), is combined with the idea of *rationality*. Some hold that these two ideas cannot be thus conjoined in discoursing about human conduct, because, they say, "causes" cannot be "reasons." It is this alleged truism, frequently asserted without

further justification, that I wish first to examine.

Let us consider a simple arithmetical example as a "pure case," one in which rational inference plays an absolutely crucial role, in which the inference is one of deductive necessity, and in which (precisely because of this deductive necessity) the behavior is determined as completely as we determine the behavior of macro-objects in ordinary physics. Let us suppose I am a practical jokester of philosophic bent. I put a jar down on the table in front of you and allow you to inspect it at leisure. I swear an oath on the Bible (let us suppose you know me to be a pious Christian believer) that I am not a magician, that there is nothing phony about the construction of the jar, and that I am not going to lie to you or engage in any kind of legerdemain. We presuppose that you take these things as true, and that you believe them with as much certainty as we can generally reach about any empirical matter. While you observe me, I now place five pennies one by one in the jar, counting out loud, and put on the lid. Then I hand you two pennies and invite you to place them also in the jar. After you have replaced the cover, I then say the following: "I want you to believe for ten seconds that there are now eight pennies in the jar. No harm will be done to anybody by your believing this, and I don't require that you assert it. So you don't even have to tell a white lie for the short run, if that would bother your conscience. I'm going, however, to attach a psychogalvanometer to your palms, and then I shall point to the figures 'six,' 'seven,' 'eight,' 'nine' on the blackboard one by one, and the instrument will reveal which numeral corresponds to your actual momentary belief about the contents of the jar. You understand I am only asking you to believe the proposition (that N = 8) for ten seconds, so you don't have to worry about developing bad arithmetical habits, or becoming psychotic through chronic reality distortions. Now, if you are able to believe for ten seconds that there are now eight pennies in the jar, I will give ten thousand dollars to your favorite charity. Surely you can have no moral objections or psychological fears about this procedure."

Now it is perfectly obvious that under these circumstances the experimental subject would very much like to entertain the proposed belief for ten seconds, but if he is a sane man, acquainted with the rules of arithmetic, it would be literally impossible for him to do so. I do not mean he would have a hard time "willing to do so," or that he would be "rationally reluctant to do so." In terms of the utilities involved, it would in fact be rational of him to make up his mind (in the pragmatic metalanguage) to believe this false sentence for ten seconds, but the fact is that it would be impossible. You could afford to wager as much money on the outcome of this experiment as you can on the outcome of a neurologist's tapping the patellar tendon to elicit the knee jerk, or on the color and size of a negative afterimage. Yet it is equally obvious that this behavior control, which is as deterministic as anyone could desire, involves a rational process, namely, the process of mental addition obeying the rules of arithmetic, as a crucial feature. By placing five pennies in the box and having the subject place 2 pennies in the box, I *determine* his belief that it now contains 7 pennies, and I render it *impossible* for him to believe that there are 8. There is, I submit, little or no more "play" in this system than there is in the elicitation of a reflex from a spinal animal or the putting of a sugar lump into solution. If the conditions stipulated are fulfilled, I

would lay as large a wager on the outcome of any one of these experiments as on any other.

There are, it seems, two opposite dangers to beware of in discussing causes and reasons in relation to human behavior. The first danger, which is not likely to be made by anyone who is philosophically sophisticated, is to conflate causes and reasons; but there is an opposite danger, one which we sometimes find in philosophically sophisticated persons, to conclude that since causes and reasons are not the same sort of entity, there cannot be any intimate connection between them, so that "explaining someone's behavior" must either be a *causal*-analysis enterprise or a *reason*-providing enterprise, but no single instance of behavior-explaining can be both. This radical separation of discourse about causes from discourse about reasons is in my view mistaken when the domain of explananda is human conduct, even though I admit (nay, insist) that the words 'cause' and 'reason' designate utterly different sorts of being. I grant the premise, that the terms 'cause' and 'reason' refer to nonoverlapping classes of designata. But I deny that from this premise we can validly infer the usual conclusion, to wit, that to provide a causal account of a person's behavior is inconsistent with giving an account in terms of his reasons. If this is paradoxical, I can only argue that it is not contradictory, and hope that its paradoxical flavor will be dissipated by sufficient immersion in my examples.

The view that I wish to develop is that while causes and reasons are utterly different sorts of things, and while in an important sense we can say that causes are "in the world" whereas reasons are not "in the world," nevertheless the *giving* of reasons, the *holding* of reasons, the *stating* of reasons, the *tokening* of reasons, the *belief* in reasons, are all psychological events, and as such are very much "in the world," and part of the chain of causality. I wish to maintain further that such psychological events have a *content*, the character of which cannot be fully set forth without employing the categories of logic. Hence, in formulating the causal laws of behavior, at least at the molar level, regarding the influencing of behavior by the tokening of reasons, the question whether or not a certain proposition or belief or sentence is a *good reason* is psychologically relevant. This is a question which can be put without conflating causes and reasons, because while a reason is not an event "in the world," the giving of a reason (or the believing of a reason, or the accepting of a reason) *is* a psychological event and *is* "in the world."

Let us take an example of simple purposive behavior to examine in this light. I mail a letter to a hotel in New York City for the purpose of arranging a room reservation because I am planning to attend a convention there. My plan to attend the convention is a good reason for sending a letter. There is a quasi-lawlike statement relating the sending of letters and the establishment in New York of a room reservation in one's name, which, while it is not a fundamental nomological, can either be made into a nomological by a suitable *ceteris paribus* clause or formulated as a statistical generalization of high p. We then have a kind of "causal law" (belonging in the domain of sociology), which relates one event, the mailing of a letter, to a subsequent event by alleging a causal connection between the two. Now this statistical generalization (or derived nomological, presupposing the *ceteris paribus* fulfilled) is known to me. And taken together with my

intention, it provides a good reason for mailing the letter. (See Hempel, 1965; von Wright, 1968; Pap, 1962, pp. 263-267. A very stimulating analysis and criticism of "the view that meaningful human actions are not amenable to causal, scientific explanation" is Brodbeck [1963]; because her mode of resolution is incommensurable with mine, and repudiates the mind-body identity thesis presupposed in Professor Popper's formulation of "Compton's problem" [see Broadbeck, 1966], I have not found it feasible under space limitations to integrate Professor Brodbeck's discussion into my paper.) In causal terms "having-the-intention-cum-believing-the-law" is a composite inner (mental) event or state that acts as an efficient cause of my letter-mailing behavior.

To say, "A reason caused my behavior" is perhaps a harmless ellipsis, but strictly speaking, it involves a confusion of the two realms which we must be careful to avoid. What we should rather say, so as to steer safely clear of this confusion, is, "The tokening of a reason was the psychological cause of my behavior." Or, lest even this formulation be taken wrongly, we could say, "The tokening of a sentence S which expresses a proposition *p*, where *p* is a good reason for action A, was the psychological cause of my emitting action A." So, even in this simple example we have at least four linkages or "connections" to consider and distinguish: First, mailing letters is a cause of room reservations expressible as a hypothetical 'If one mails a letter, he gets a room'; second, this causal relation is, *in the realm of inference*, a good reason for mailing a letter, granted the premise that one wants a room reservation; third, my tokening of this good reason functions as a psychological cause of my performing an act whose description occurs in the antecedent statement of the hypothetical (a relation in pragmatics which, in general, is a characteristic of purposive behavior); fourth, an external observer would in turn have a good reason for expecting me to mail this letter, and that good reason would be *his* understanding of the psychological causal law which says that, *ceteris paribus*, if a rational person wills the consequent of a causal law which the person believes, he tends also to will the antecedent. (The *ceteris paribus* clause must, of course, include such qualifiers as 'absent countervailing means-end structures' and the like.) In this analysis I have not, I trust, anywhere conflated causes with reasons. Yet I have explicitly recognized that a critical element in what makes certain kinds of mental events causally efficacious is that they are tokenings of sentences which, in the realm of logic, constitute valid reasons.

Consider next the case of a simple desk calculator. In order for it to compute sums accurately, its internal structure must have some kind of isomorphism with decimal arithmetic. Thus, the machine is constructed so that after a wheel has turned through ten positions, this physical fact causally produces a one-position displacement in the next adjacent wheel, i.e., the wheel which "corresponds" to the next integer to the left. The machine behaves rationally, in that it makes legitimate or valid transitions in the arithmetical language game. If it were not constructed in the way it is, or in some alternative way preserving the necessary machine-arithmetic correspondence, it would not be able to do this. We telephone the company and ask for a repairman to be sent out when the machine begins to make counter-arithmetical transitions, i.e., it "makes mistakes" and "gives the

wrong answers." (Even a philosophy professor normally finds these locutions quite natural under such circumstances.) Such a desk calculator is clearly a "clock" rather than a "cloud," but as it gets old and worn out and becomes a little more cloudlike, it also becomes more irrational, i.e., slippage in the gears leads it to make arithmetical mistakes. We can carry the analogy further, still remaining at the level of a mere desk calculator rather than the big modern computers. What the machine will do with the numbers we punch in depends upon our giving it instructions, which is (formally) comparable to the "intention" or "mental set" adopted by a human being as he listens to us giving reasons. It is no objection to this analogy that the machine does not have conscious intentions, because it is imperative to distinguish the components of sentience and sapience (Feigl, 1967; Feigl & Meehl, [1974]; Meehl, 1966) and what we are concerned with in the present section is whether determinism is in any way incompatible with that aspect of *sapience* which we call 'rationality.' (Even in the human being there is, of course, plenty of evidence to say that sapience can occur, and sometimes in very complicated forms, in the absence of reportable phenomenal events. The well-known examples of unconscious literary composition or scientific problem-solving, not to mention the quite complicated content of means-end connections involved in psychoanalytic mechanisms, suffice to show this.)

With the kind permission of Professors Schilpp and Freeman, I would like to quote here a passage from the forthcoming contribution by Professor Feigl and myself to the Schilpp volume on Sir Karl Popper's philosophy.

> Returning to the question of the sense in which a physicalistic account in brain-language is "complete" *even though it does not say all that could be said*, we suggest the following as a first approximation to an account which, while maintaining the distinction between logical categories and the categories of physics or physiology, nevertheless insists that a physicalistic micro-account is nomologically complete. We have a calculus, such as arithmetic or the rules of the categorical syllogism. We have a class of brain-events which are identified by appropriate physical properties—these, of course, may be highly "configural" in character—at, say, an intermediate level of molarity (i.e., the events involve less than the whole brain or some molar feature of the whole acting and thinking person, but are at a "higher" level in the hierarchy of physical subsystems than, say, the discharge of a single neuron, or the alteration of microstructure at a synapse). Considered in their functioning as inner tokenings—that is, however peripherally or behavioristically they were originally acquired by social conditioning, considering them as now playing the role of Sellars' *mental word* (Sellars, 1956, 1966; Chisholm & Sellars, 1958)—there is a physically-identifiable brain-event b_M which "corresponds" (in the mental word sense) to the subject-term in the first premise of a syllogism in Barbara. There is a second tokening event b_P which is a token of the type that designates the predicate-term of the conclusion; a brain-event b_S which corresponds to a tokening of the type that designates the subject-term of the conclusion of the syllogism; and finally a brain-event b_C corresponding to the copula. (These expository remarks are offered with pedagogic intent only. We do not underestimate the terrible complexity of adequately explaining the words 'correspond' and 'designate' in the immediately preceding text.)
>
> A physically-omniscient neurophysiologist [= Omniscient Jones estopped from meta-talk about logic] can, we assume, identify these four brain-events b_M, b_P, b_S, b_C on the basis of their respective conjunctions of physical properties, which presumably

are some combination of *locus* (where in the brain? which cell assemblies?) and *quantitative properties of function* (peak level of activation of an assenibly, decay rate, pulse-frequency of driving the next assembly in a causal chain, mean number of activated elements participating). For present purposes we may neglect any problem of extensional vagueness, which is not relevant to the present line of argument, although it is of considerable interest in its own right.

Our physically-omniscient neurophysiologist is in possession of a finite set of statements which are the nomologicals (or quasi-nomologicals) of neurophysiology, which we shall designate collectively by L_{phys} [= neurophysiological laws]. He is also in possession of a very large, unwieldy, but finite set of statements about structure, including (a) macrostructure, (b) structure of intermediate levels, e.g., architectonics and cell-type areas such as studied microscopically in a brain-histology course, and (c) micro-structural statements including micro-structural statements about functional connections. We take it for granted that "learned functional connections" *must* be embodied in micro-structure (although its exact nature is still a matter for research) since there is otherwise no explanation of the continuity of memory when organisms, human or animal, are put into such deep anesthesia that all nerve cell discharge is totally suspended for considerable time periods, or when normal functional activity is dramatically interrupted by such a cerebral storm as a grand mal seizure induced in electroshock treatment. Thus the class of structural statements S_t includes two major sub-classes of statements, one being about the inherited "wiring diagram" of a human brain, and the other being the acquired functional synaptic connections resulting from the learning process.

Our omniscient neurophysiologist can derive, from the conjunction (L_{phys} . S_t), a "brain-theorem" T_b, which, to an approximation adequate for present purposes, may be put this way: Brain-state theorem T_b: "Whenever the composite brain events ($b_M b_C b_P$) and ($b_S b_C b_M$) are temporally contiguous, a brain-event ($b_S b_C b_P$) follows immediately." This brain-theorem is formulated solely in terms of the states b_i which are physicalistically identifiable, and without reference to any such meta-concept as class, syllogism, inference, or the like. The derivation of T_b is one of strict deducibility in the object-language of neurophysiology. That is, neurophysiology tells us that a brain initially wired in such-and-such a way, and then subsequently "programmed" by social learning to have such-and-such functional connections (dispositions), will necessarily [nomological necessity] undergo the event ($b_S b_C b_P$) whenever it has just previously undergone the events ($b_M b_C b_P$) and ($b_S b_C b_M$) in close temporal contiguity.

But while for the neurophysiologist this brain-theorem is a theorem about certain physical events *and nothing more,* a logician would surely discern an interesting formal feature revealed in the descriptive notation—the subscripts—of the *b*'s. It would hardly require the intellectual powers of a Carnap or Goedel to notice, *qua* logician, that these brain-events constitute a physical model of a sub-calculus of logic, i.e., that these physical entities [b_M,b_P,b_S,b_C] "satisfy" the formal structure of the syllogism in Barbara, if we interpret

b_M = tokening of middle term b_S = tokening of subject term
b_P = tokening of predicate term b_C = tokening of copula

The "brain-theorem" T_b can be *derived nomologically* from the structural statements S_t together with the microphysiological law-set L_{phys}, given *explicit definitions* of the events [b_M,b_P,b_S,b_C]. These explicit definitions are not the model-interpretations, nor are they "psycholinguistic" characterizations. We can identify a case of b_P by its physical micro-properties, *without knowing* that it is a tokening-event, i.e., without knowing that it plays a certain role in the linguistic system which the individual who

owns this brain has socially acquired. But brain-theorem T_b has itself a *formal structure,* which is "shown forth" in one way, namely, by the syntactical configuration of the b-subscripts [M,P,S,C]. In this notation, "which subscript goes with what" is determinable, so long as the events b_i are physically identifiable. There is nothing physically arbitrary in this, and there is nothing in it that requires the physically-omniscient neurophysiologist to be thinking about syllogisms, or even, for that matter, to know that there is any such thing as a syllogism. Although again, it goes without saying that he himself must reason logically in order to derive the brain-theorem. But he does not have to meta-talk about rules, or about his own rule-obedience, in order to token rule-conformably in his scientific object-language, and this suffices to derive T_b.

One near-literal metaphor which we find helpful in conveying the essence of the "syllogistic brain-theorem" situation, as we see it, is that the sequence of brain-events $(b_ib_jb_k)$ $(b_jb_kb_l)$ … *embodies* the syllogistic rules. Their defined physical structure plus the physical laws of brain function causally necessitate that they exemplify syllogistic transitions, a fact revealed when the notation designating them is considered in its formal aspects. In the usual terminology of thinking processes and logic, the brain-theorem T_b says, in effect, that the existence of a formal relation of deducibility (truth of logic) provides, in a brain for which the theorem obtains, the necessary and sufficient causal condition for a factual transition of *inference* (a mental process). This assertion may appear to "mix the languages," to "commit the sin of psychologism," to "conflate causes with reasons"; but we maintain that none of these blunders is involved. It is a *physical* fact that a certain *formal* relation is physically embodied. If the formal features of the initial physical state were otherwise, the ensuing physical result would have been otherwise. Hence the physical embodiment of the formal relation—a *fact*, which is "in the world" as concretely as the height, in metres, of Mount Everest—is literally a condition for the inference to occur. [Feigl & Meehl, 1974, pp. 548-550]

I need hardly say that the idea that strict rationality in a deductive-inference situation is not only compatible with determinism but at the common-sense level requires it—"If I am 100% rational, I will be *unable* to deny conclusions strictly implied by premises"—is hardly a new insight on my part, and I have not felt it useful to canvass the philosophical or psychological literature for citations. Since the first draft of this paper was written, two explicit statements on this point have been brought to my attention, one by Ruth Macklin (1968; see also Macklin, 1969), in an illuminating paper entitled "Doing and Happening," where we read:

The problem of trying to make this distinction [between things a person does and things that happen to him] hold for all cases becomes even more complex when we consider mental acts such as believing, thinking, and wanting. Although choosing, deciding, and forming intentions appear to be mental acts in the sense that they seem to be clear cases of something a person does, what about believing? Does a person choose to believe the things he believes? Or to think the thoughts he thinks? Does he have control over his beliefs in the psychological sense that he can, in fact, avoid believing that *p* in cases where evidence in favor of the truth of *p* is overwhelming? If he cannot control his beliefs in such cases, are we to say that believing that *p* is not something which that person does, but rather something that happens to him? This result is obtained by using an analogue of the physiological control criterion which may be somewhat infelicitously termed "the mental control criterion." It does seem counter-intuitive to claim that believing is not something that someone does; yet it is

> not clear that either the mental control criterion or another appeal to linguistic usage will answer the question satisfactorily. We do sometimes say, "I cannot help believing that," or "Try as I might, I cannot believe that," indicating that the ability to choose or control our beliefs is open to question. This problem can be met, in part, by making the further distinction between deliberate and non-deliberate doings, and between believings that are reflective and those that are not. Hence, application of the mental control criterion would result in the position that some types of believing are not things that one does, but rather things that happen to one. Perhaps, then, the criterion should be rejected. But on what grounds? Presumably, on the grounds that it conflicts with our intuition that believing is always something persons do. Of course, there is still another alternative, namely, that the distinction between what a person does and what happens to him is inapplicable to mental acts such as believing. On this view, it is inappropriate to claim that believing is either something that one does or something that happens to him. [Macklin, 1968, pp. 257-258]

The other quotation, as succinct and explicit a statement of my position as one could easily find anywhere, goes back to 1905, in Max Weber's critique of Eduard Meyer's methodological views.

> The error in the assumption that any freedom of the will—however it is understood—is identical with the "irrationality" of action, or that the latter is conditioned by the former, is quite obvious. The characteristic of "incalculability," equally great but not greater than that of "blind forces of nature," is the privilege of—the insane. On the other hand, we associate the highest measure of an empirical "feeling of freedom" with those actions which we are conscious of performing rationally—i.e., *in the absence of physical and psychic "coercion," emotional "affects" and "accidental" disturbances of the clarity of judgment,* in which we pursue a clearly perceived end by "means" which are the most adequate in accordance with the extent of our knowledge, i.e., in accordance with empirical *rules*. If history had only to deal with such rational actions which are "free" in this sense, its task would be immeasurably lightened: the goal, the "motive," the "maxims" of the actor would be unambiguously derivable from the means applied and all the irrationalities which constitute the "personal" element in *conduct* would be excluded. Since all strictly teleologically (purposefully) occurring actions involve applications of empirical rules, which tell what the appropriate "means" to ends are, history would be nothing but the applications of those rules. The impossibility of purely pragmatic history is determined by the fact that the action of men is *not* interpretable in such purely rational terms, that not only irrational "prejudices," errors in thinking and factual errors but also "temperament," "moods" and "affects" disturb his freedom—in brief, that his action too—to very different degrees—partakes of the empirical "meaninglessness" of "natural change." Action *shares* this kind of "irrationality" with every natural event, and when the historian in the interpretation of historical interconnections speaks of the "irrationality" of human action as a disturbing factor, he is comparing historical-empirical action not with the phenomena of nature but with the ideal of a purely rational, i.e., absolutely purposeful, action which is also absolutely oriented towards the adequate means. [Weber, 1949; I am indebted to my Law School colleague Professor Carl Auerbach for calling this reference to my attention.]

I have no doubt made my task somewhat easier, as Sir Karl might object, by confining myself to that restricted form of rationality involved in thinking syllogistically. But while the choice of such an example simplifies the problem, I

cannot think that the use of such an example is tendentious or prejudicial. And this is the more so since Sir Karl, whose position is under examination here, is such a firm and articulate opponent of the idea that there exists such a thing as "inductive logic." Without making overmuch of the Reichenbachian dichotomy between the context of discovery and the context of justification (a dichotomy which is not clear-cut, but still useful for many purposes) I would suggest that those portions of what is ordinarily considered subsumable under "inductive inference" that involve the testing of hypotheses and generalizations can be given a syllogistic form (e.g., *modus tollens*), so that the preceding syllogistic example can function as a paradigm case for dealing with Sir Karl's position in this respect; and that what is not so formulated can *either* (a) be viewed as satisfying or not satisfying some overarching methodological prescription in the pragmatic metalanguage (and hence examinable in an essentially syllogistic way, i.e., we inquire whether a given bit of concrete pragmatics of inference is or is not in accord with the methodological prescription), or (b) really properly relegated to "context of discovery" in the strong sense of the phrase. Since Sir Karl himself explicitly repudiates the problem of the psychology of discovery as not belonging to the logic of the matter, the question how (historically, psychogenetically, sociologically) a particular scientist comes to hit upon a theory or an experimental arrangement is not relevant to our present issue. The possibility of novelty, of genuine "creation" by the scientist or artist, does involve this context-of-discovery question, and Sir Karl adduces it as a further objection to determinism; but that is outside the scope of this paper (see, however, Feigl & Meehl [1974]).

One's philosophical (and perhaps, more importantly, one's "personal, human, existential") discomfort about determinism in relation to the possibility and limits of human rationality is, I suggest, often exacerbated by two mental habits in our thinking about the psychological causation of beliefs and related cognitive processes (e.g., perceiving, inferring, recalling). The first is our habit of associating the idea of psychological causation primarily with the nonrational or irrational class of causes, such as unquestioned beliefs carried over from childhood, political manipulation of mass opinion through propaganda, personal idiosyncratic prejudices having a variety of origins, unconscious determinations of the Freudian type, and the like. One has the impression, in talking with either philosophers or "plain men," that when asked to contemplate the possibility that their own behavior may be completely (or even largely) determined by antecedent causal conditions, they tend to think immediately of such factors as the fried eggs they had for breakfast, or the prejudice they learned from their Norwegian grandmother against the Danes, or the subtle influence of TV advertising upon consumer choice, rather than such psychological facts as the fact that they have been presented with certain evidence in the form of statements from reliable sources, or have been subjected to criticism in the course of discussion with a colleague, or have made certain observations in the laboratory. Thus it seems that our tendency to polarize "human reason" over against "psychological causality" infects our thinking by influencing the very examples that occur to us in this connection, so that we tend to think of psychological causality solely in terms of the kinds of causes which normally are used to explain a piece of human irrationality. No

doubt the influence of Freud and Marx, at least among educated persons, is important here, inasmuch as they both stressed the "hidden, nonrational" forces that play a greater role in the molding of man's opinions than had been formerly supposed.

The second habit is our tendency to connect such "psychological determiners" as motives, affects, training, group pressures, psychoanalytic identifications, and the like almost entirely with a person's *particular substantive* (*object-language*) *opinion*, forgetting to take into account the fact that these same classes of psychological determiners also exert a powerful causal influence upon his *meta*-talk and *meta*-thought. That is, we readily recognize the possibility that I am a zealous pacifist or jingoist because I have a strong unconscious father-identification and my father was a pacifist or jingoist, as the case may be. But we neglect the possibility (thankfully, one realized in at least an appreciable minority of human beings!) that I am committed to certain overarching procedural or methodological principles, such as rationality, critical discussion, and the examination of contrary evidence, as a result of my psychological history. These overarching methodological habits are, of course, subject to learning by reward or father-identification or peer-group conformity pressures, and it is a mistake to assume that only our specific-issue opinions are "psychologically caused." In common life, we have occasion to characterize persons according to their meta-talk dispositions. We may say of a certain individual, "Well, he's usually a very rational fellow on most subjects, and he tries hard to be fair-minded and to see the other fellow's viewpoint; but you'd better not get him on the race question, because there he goes haywire." What do we mean by a remark of this kind? We mean that this person is one who has developed very strong pervasive "meta-habits" of the kind we call 'rational,' so that we expect that these overarching considerations, e.g., his self-concept of being a rational person, and his sincere desire to find out the truth about matters to which he addresses himself, his compulsion to derive implications from hypotheses and compare them with facts, will *in general* be controlling with respect to his processes of thinking about particular substantive matters; but that this overarching monitor or control system, which causes him to think rationally in general, is not sufficiently strong to countervail the influence of a very strong emotional commitment on this particular substantive issue of the race question. (Each of us has on occasion in his smaller way to make Henry Clay's famous decision on whether we would rather be right than be President!)

It may be objected that in these remarks I have fallen into the confusion I earlier renounced, to wit, mixing the category of "causes" with the category of "reasons." If so, it has been through some subtle philosophical mistake, because in writing the foregoing I had this distinction constantly in mind. But I have permitted myself such locutions as "thinking rationally," or "countervailing influence against rationality," and it is obligatory upon me to explicate the cause-reason relationship indicated by such language.

What, then, is the situation here, as regards the relation of causes to reasons, when we inquire concerning a particular individual's thinking about a certain subject whether it was mainly influenced by "rational" or "irrational" factors? For ease of exposition and to avoid nonrelevant issues in psycholinguistics (but, I

hope, without loss of generality), I confine myself to thinking processes sufficiently "symbolic-linguistic" in nature to make appropriate the notion that the individual tokens a sentence. I do not mean thereby to prejudge the question whether all forms of intentionality involve sentence-tokening; but that sort of intentionality which is (debatably) present when I simply image a state of affairs, or when I have an unworded "expectation" that is rudely disappointed (as in Russell's well-known example of a nonlinguistic belief, that of experiencing surprise upon finding another step on the stairs when I thought I had come to the bottom even though I was not consciously "thinking about it" at all) is too dubiously "rational" to be of use for our analysis. I have no stake in asserting that such non-worded, inchoate expectations *cannot* (on some suitable reconstruction) be considered rational; and in fact I tend to believe some of them should be called so. But since the detailed reconstruction is not available, and since some readers would disagree with me, I avoid these marginal cases. The issue is the compatibility of rationality and determinism, and it seems unlikely that the alleged incompatibility would show up in the case of dubiously intentional (mental) acts of a nonsymbolic, nonlinguistic sort, but for some strange reason be lacking in the clear-cut linguistic case. And since the linguistic case is the one which has been subjected to more adequate philosophical analysis, it is the only one I shall consider here.[2]

Consider an example in which I begin without any personal involvement or emotional prejudice one way or the other. I am an educated man but a non-mathematician, and I have no particular philosophical leanings about the idea of infinity. In a semipopular volume on mathematics, I come across the question whether there is a largest prime. I think to myself, "That's an interesting question; it never occurred to me; I shouldn't be surprised one way or the other; and I couldn't care less." I read through Euclid's short and easy proof, which I find convincing. From that moment onward, I firmly believe that there is no largest prime. This would seem to be a rather clear-cut case of practically 100 percent rationality. Now in what sense, if any, can we speak of the "causes of my belief" in the infinity of primes being "rational causes," without conflating the distinct categories of *cause* and *reason*?

It seems to me that there is no great mystery here, that the correct analysis is quite straightforward, and not even paradoxical. In reading through the proof, I token consecutively the sentences which, according to the syntactical rules of the language I speak, constitute (in the realm of logic) a formally valid proof of the infinity of primes. These sentences express propositions which, in the realm of logic, constitute "valid grounds for believing the conclusion." That is, the propositions which these sentences express are *good reasons.* What makes them good (deductive) reasons is, of course, that the conclusion can be reached by a finite number of steps, each step being taken in accordance with the transformation rules of the language. The tokenings (= my *thinking the sentences*) are not, strictly speaking, reasons. The tokenings are psychological causes, that is to say, they are events which go on in my mind (or, if physicalism is true, we can also

[2] For a fascinating—although, in the end, somehow unsatisfying—analysis of the concept rational as applied to imaginary dance-language of bees, see Bennett (1964).

say "in my brain") (see Meehl, 1966, pp. 108, 160-163). What makes them effective causes is the fact that my brain is wired (or, more accurately, we would have to say wired plus programmed) to make language transitions in accordance with certain syntactical transformation rules. When I token a sentence S_1 and a sentence S_2, tokenings of sentences related such that the sentence S_3 logically follows from S_1 and S_2 in accordance with the transformation rules of the language in which I have learned to "think," I am strongly disposed (to the extent that I am a rational man) to token S_3. What is the mystery about this? Except for vast differences in complexity, how does this differ from the fact that my desk calculator has a mechanical construction such that the movements of its gears "obey the laws of arithmetic"? Elliptically, it is therefore unobjectionable, provided there is no danger of confusion, to say that I am "caused to believe in the infinity of primes by valid reasons." But this locution should probably be avoided in the interests of clarity. The reasons are not causes, but the tokenings of the sentences which express the reasons are causes. If my terminal tokening of Euclid's conclusion is in fact psychologically produced not by admiration of Euclid (for some students the letters Q.E.D. could stand for "quod Euclid dixit"), or by the fried eggs I had for breakfast, but by the fact that my brain has been programmed to token sentences in accordance with transformation rules, then my belief in the infinity of primes is "rationally determined."

I have in this analysis made free use of the notion of an inner "tokening" as if it were an obvious and clear idea, which it admittedly is not. Unfortunately this is one of those concepts in whose consideration philosophy unavoidably overlaps with one of the empirical sciences, namely, psycholinguistics; and psycholinguistics is a science presently in a primitive state of development so that it cannot provide clear-cut, well-established laws (or, hence, adequate implicit definitions of its theoretical entities) for the use of the philosopher interested in semantics. It seems clear that the mental word need not possess a complex internal structure capable of correspondence, picturing, or "isomorphism" with the external nonlinguistic event that it designates. For example, a single noise, not capable of division into parts or components which have a separate "meaning," may, to an Eskimo, be equivalent to an English language statement, "There is today a great deal of snow, of a slushy variety." No psycholinguist or philosopher can, in the present state of knowledge, specify in detail what are the necessary and sufficient physical and psychological conditions for an Eskimo to token this word internally. Fortunately, it is not necessary for the philosopher to rely heavily upon technical psycholinguistics in discussing Popper's thesis. All that we need suppose is that there occurs a certain kind of physical event in the brain, whatever its "internal" nature, that has the required relation to an external event such that it is entitled to be called a tokening of a sentence, when that sentence expresses a proposition designating the external event. In the case of an English speaker, the resources of his language are such that he must say "There is a great deal of snow, of a slushy type." Whereas for an Eskimo speaker it may be sufficient to say "glop," which means the same thing to him as the more complicated expression means to an English speaker. The point for our purposes is that *whatever* physical event in the brain has come (by social learning) to possess this kind of statistical

correspondence to slushy snow, perceptions normally produced by slushy snow, expectations about correlates of slushy snow, etc., constitutes the physical tokening of the proposition. The occurrence of slushy snow is physically describable. The occurrence of a particular token event is also physically describable. (I set aside ontological dualism for the time being; but since determinism and not materialism is in issue, I do not believe that prejudices our present discussion. Would not Popper's objection hold against a deterministic dualistic interactionism, if it holds against a deterministic identity thesis?) Examining the role of a given kind of tokening event in the total tokening system of an individual belonging to a particular culture, we can (in principle) determine the external event to which it "refers" (insofar as it refers precisely, which it almost never does). Having determined that, we understand the "meaning" of the sentence which the individual tokens. And if he tokens that sentence when the external event does not occur, then we say that he "tokens falsely."

Of course we know that the formulation of semantic rules on the basis of studying a natural language ("English as she is spoke") will always involve a certain element of arbitrariness, because we may or may not choose to embody certain statistically deviant locutions in the rules. This will depend upon their statistical rarity, in large part, although not wholly. (Cf. the dictionary-maker's problem of deciding between "second usage" and "erroneous usage.") Thus Carnap (1939, pp. 6-7) says that after observing the verbal behavior of people who speak a particular language B and noticing that 98 percent of the time they use the word 'mond' to mean the moon, but 2 percent of the time use 'mond' to refer to a kind of lantern, we are free to formulate the semantics of language B either to include this special and rare usage or not. That is, the formulation of a pure semantics, like any idealization or rational reconstruction, is something done with *attention to* (and *on the basis of*) the empirical statistics of descriptive semantics and pragmatics, but cannot usefully aim to *exactly correspond* to the latter. (If it did, as a desideratum, it would not be rule-construction, since a rule is, by definition, something capable of being violated.)

In the present state of knowledge of how the human brain mediates behavior and experience (on *either* a dualistic or "identity" view of the mind-body problem) it is pointless to speculate philosophically about possibilities as regards detail. Nor can one anticipate with any confidence just where on the cloud-clock continuum various kinds of psychological processes will be located when psychophysiology has reached a relatively advanced state. When the contemporary psychologist deals with rational behavior, such as that of a logician performing the task of classifying a simple syllogism as formally invalid, he proceeds in the same way as the philosopher or the layman does, namely, at what is usually called the "molar" level of analysis (Tolman, 1932, pp. 3-23 and *passim*, 1951; Hull, 1943, pp. 19-21; Skinner, 1938, pp. 3-6, 33-43; Littman & Rosen, 1950; Murray, 1938, pp. 55-58, 96-97; MacCorquodale & Meehl, especially pp. 218-231; and the operant behaviorists generally, cited in footnote 1 *supra*). That is to say, in forecasting Professor Popper's verbal response to a syllogism which commits an Illicit Distribution of the Major, we do not—in part because we *cannot*, at the present state of knowledge of brain function—mediate this prediction in the

language of neurophysiology. Instead we rely upon dispositional properties (setting aside for the moment whether these are probabilistic or nomological) of the "whole organism," the man Popper, whom we know to be sane, sober, attentive to his task, and who has a history of having been educated in formal logic. The fact that our predictions of his verbal behavior are mediated on this molar basis gives rise to an interesting problem in the methodology of those sciences that deal with the behavior of human beings, namely, to what extent must the behavior scientist employ the concepts of the logician?

It is important to make some distinctions here which are sometimes overlooked. A logical category, such as "valid syllogism," can be involved in the psychologist's task in three different ways. First, the psychologist wants to proceed rationally in his own scientific thinking, i.e., it is necessary that he himself as a knowing organism *exemplify* or *be obedient to* the laws of logic. That is to say, in his own (object-language) discourse he must avoid committing fallacies. Second, we recognize that a considerable amount of scientific writing and discussion involves, in addition to object-linguistic assertions describing observations or propounding theories, processes of rational criticism. Here the psychologist moves periodically into the metalanguage as he engages in such processes as evaluating experimental evidence, examining the theoretical derivations offered by himself and others, carrying on rational inquiry about the internal consistency of a system of theoretical propositions, and the like. So far as I can discern, in these two respects the behavioral scientist's "use" of logic does not differ in any essential way from that of the botanist or the astronomer. All scientists must think logically, whether in the object language (substantive derivations and classifications) or in the metalanguage (criticism, evaluation, and research strategy). But it seems that insofar as the psychologist treats of human cognitive processes at the molar rather than at the neurophysiological level, he is forced to employ the concepts of logic in a third way, a way that is unique to social science as a subject matter. The reason for this peculiarity of psychology, sociology, economics, etc., as Professor Popper would be the first to emphasize, is, put most simply and directly, that plants and stars do not think, but human beings do.

This undisputed fact (did even John B. Watson *really* doubt it?) about the special nature of the psychologist's subject matter gives rise to the paradox that *concepts customarily regarded as metalinguistic unavoidably appear in the psychologist's object-linguistic discourse whenever he is attempting to mediate predictions about rational human behavior.* It might be supposed that more adequate behavioristic formulation of rational behavior could, in principle, dispense with the employment of such metalinguistic concepts. Quite apart from the fact that this hope refers to a Utopian state of behaviorism, whereas we all admit that the psychologist, like the layman or the philosopher, can successfully mediate high-probability predictions given the present *non*-Utopian state of the psychology of cognition and psycholinguistics, I must further point out that it is far from obvious that even in a Utopian state of these molar disciplines it would be theoretically possible to dispense with the logician's metalanguage in giving a psychological description or causal analysis of rational human behavior. Since the possibility of their permanent indispensability at the molar level of analysis is the alternative

most favorable to Professor Popper's position, let us scrutinize the consequences of this alternative in more detail.

Consider again the simplified, idealized example of a logician being confronted with a syllogistic argument containing an Illicit Major. We set aside the empirical problems involved in ascertaining those aspects of the individual's emotional and motivational state which are relevant to his momentary disposition to think rationally. That is, we assume that the "test conditions" for activating his disposition to classify an argument as Illicit Major are known to be momentarily fulfilled. The usual gambit for an arch-behaviorist who aims at eliminating any vestige of "mentalistic concepts" from his descriptive or theoretical language is to invoke the individual's learning history and to infer from it that the logician has a very strong "habit" of responding with the phrase "Illicit Major" when he is presented with a certain stimulus, say, the formally invalid syllogism appearing in three lines on a clearly printed page. Leaving aside the current controversies in psycholinguistics (which call into serious question the theoretical adequacy of any "stimulus-response" model of verbal processes) let us proceed on the (probably false) assumption that such an analysis could be given and satisfactorily corroborated. That is, let us assume that the molar behaviorist can make good on his claim of accounting for the logician's current disposition to token the metalinguistic expression "Illicit Major" as a response to the printed tokens of such a formally invalid syllogism. The question now arises, how is the stimulus class to be characterized in molar language? As is well known, there are terrible difficulties involved in the whole problem of pattern recognition, such that no one has as yet provided an adequate theoretical model of the necessary structures which, if we were in possession of it, would enable us to construct and program a computer to duplicate even fairly simple visual pattern-recognition functions of the human brain (Sayre, 1965). We shall set this whole class of difficulties aside also, and assume as an oversimplified situation that the typeface, size, spacing, etc., are physically identical with those which have been presented to the logician in his previous experiential history.

But even these idealizations and oversimplifications do not, it seems to me, get rid of the behaviorist's fundamental problem. We know that it is possible to present the logician with *any* syllogism having the requisite formal structure of an Illicit Major, and confidently predict that his response will be a tokening of 'Illicit Major,' or some equivalent thereof. (Let us set aside the problem of what are "equivalent responses" and confine our attention solely to the problem of identifying the stimulus class.) Now the fact that the presented visual stimulus, a syllogism on the printed page, need not be physically identical in terms of mounds of ink with any stimulus previously presented to the logician is not in itself a serious objection to the behaviorist's analysis, it being admitted on all sides that some underlying concept of stimulus equivalence or stimulus generalization will be required by any adequate molar theory (since from the sheer standpoint of their physics no two stimulus inputs, at least among those occurring in ordinary life, are strictly identical). That is, what Skinner in *The Behavior of Organisms* calls "the generic nature of the concepts of stimulus and response" is taken for granted by psychological theorists of many different persuasions. The scientific problem

here is not (at the molar level) to derive or explain the basic phenomenon of stimulus equivalence or stimulus generalization, which is rather taken as a rock-bottom fact, a basic postulate in any molar behavior theory, and presumably finds its own explanation in turn at another level of causal analysis, i.e., at the neurophysiological level. The problem at the molar level, once having included some suitable theoretical postulate regarding stimulus equivalence or stimulus-generalization gradients, is that of *formulating*, in the descriptive language which we employ to characterize the stimulus side, *what the common property of the stimulus inputs which belong to a stimulus-equivalent class must be*. Or, speaking not in terms of strict stimulus equivalence but rather in terms of the stimulus-generalization gradient, the problem is one of formulating the relevant features of the physical dimensions which constitute the input variables with respect to which the generalization gradient is to be plotted as a hyper-surface in a stimulus hyper-space. To avoid the mathematical complexities involved here, we shall simplify further by speaking of stimulus equivalence rather than stimulus-generalization gradients, i.e., we shall dichotomize the syllogism inputs into illicit and licit distributions. We shall also neglect the fact that even for a logician there might be certain formal presentations which would be more "seductive" in leading him to misclassify the syllogism as valid when it is actually fallacious (Woodworth & Sells, 1935; for a summary of this and related studies, see generally Woodworth, 1938, pp. 810-817). Our problem then becomes, how do we characterize the stimulus input in a molar-psychological formulation of the stimulus side, of a verbal habit whose response side consists of the tokening 'Illicit Major'?

Now it is a truism, routinely pointed out to students in an elementary logic course, that the fallacious character of such a syllogism is revealed by its form alone, so that one can identify an Illicit Major even if the terms (other than the logical constants) are terms whose meaning is not known to the classifier. For that matter they could be neologisms which have no meaning, in anybody's natural or artificial language. So that when we present to the logician a syllogism which says, "No glops are klunks; all klunks are fabs; ergo, no glops are fabs," we will be perfectly confident that he can respond with the tokening of 'Illicit Major' in spite of the fact that the terms 'glop,' 'klunk,' and 'fab' are novel to him. And we have this confidence because we know that, for a logician, the defining property of an Illicit Major is its possession of a certain syntactical form, i.e., all syllogisms of this form are stimulus-equivalent to him as determiners of the verbal response 'Illicit Major.' The possession of this syntactical form is both a necessary and a sufficient condition for the logician to token the fallacy's name.

It might be argued by the staunch behaviorist that he can describe the syllogism-input stimulus class without making use of the logician's concept *Illicit Major*. Now no one wants to maintain that the behaviorist must employ the logician's *terminology*, i.e., he need not employ the actual expression 'Illicit Major' to mediate his prediction. But does this get around the behaviorist's difficulty? I do not think it does. After all, the metalinguistic expression 'Illicit Major' is introduced by the logician through explicit definition and, therefore, is in principle eliminable from *his* discourse as well. But the point is that when the behaviorist attempts to really deliver the goods on his claim to be able to

characterize the stimulus side of the logician's disposition, he will find himself unavoidably driven to set forth the formal (syntactical) characteristics of the adequate stimulus class; and these characteristics will (if the logician is really a logical logician!) be identical with the defining syntactical property which the logician expresses by the shorthand phrase 'Illicit Major.'

It would therefore seem more honest for the behaviorist to admit from the start that he employs the syntactical category referred to by the metalinguistic expression 'Illicit Major,' and that being the case, he might just as well include the logician's phrase in his scientific vocabulary and be prepared to utilize it in object-language derivations.

Should this distress him? I think not. Consider an analogous infra-human example. An animal psychologist is studying form discrimination in the monkey, say in a Skinner-box situation, where the discriminative stimulus is the presentation of a visual pattern on an illuminated screen and the response alternatives are to depress one of two levers. During the training phase of the experiment, the monkey is reinforced if he presses the right-hand lever in the presence of an isosceles triangle composed of three straight lines, and he is also reinforced for pressing the left-hand lever if the visual stimulus is a single circle of approximately the same size as the triangle. After this discriminative control is thoroughly established, the experimenter then presents the monkey with a visual pattern consisting of three small circles in a triangular arrangement and lacking symmetry of placement such that the circle which constitutes the top vertex is displaced leftward from the midline between the other two. So the novel visual stimulus on test trials consists not of an isosceles triangle formed out of three visible straight-line segments, as in training trials, but rather of the three vertices only, arranged so that they define a scalene triangle, and the vertices are circles which, although of smaller size, are, as circles, members of the same geometrical class as the original training stimulus for pressing the *left*-hand lever. It is likely that if these circles were made very large and close to collinear, the monkey would respond to them as approximately stimulus-equivalent to the original circle; if they are made very small, almost points, he may or may not respond; and with a considerable departure from collinearity he will (one hopes) respond to them as triangular, in spite of the fact that the physical lines connecting the vertices of this triangle are missing.

Suppose the psychologist, by trying various combinations after such original training, finds that the probability or strength of the response disposition to the right versus left lever depends in a complicated way upon (a) the absolute size of the original negative circle, (b) the absolute sizes of the test circles used as vertices, (c) the ratio between these circular areas, (d) the degree of departure from collinearity of the three points in the test trial, (e) the distances between the vertices in relation to the angles of the test triangle, and so forth. Let us imagine that (whether on theoretical grounds or by a blind, curve-fitting process) the investigator succeeds in constructing a complicated configural function which relates the response strengths of the two lever-pressing operants to these geometrical features. That is, he writes response-strength equations or probability equations of the type $P_R = F(L_1, L_2, R_1, R_2, \theta)$ and $P_L = G(L_1, L_2, R_1, R_2, \theta)$.

Obviously, in explaining what these variables are, our psychologist employs concepts of analytic geometry and trigonometry, i.e., he has to explain that the variable θ which appears in these response-strength functions refers, say, to the smaller of the two angles between the lines connecting the vertices of the circles on the test trial, and so forth. Thus he employs, in his descriptive discourse characterizing the stimulus side, an interpreted formalism, i.e., physical geometry.

Now suppose either a hard-nosed behaviorist or an anti-behaviorist philosopher with intent to gore the behaviorist's ox were to object to this procedure by saying, "But, my dear fellow, you said that you were a behaviorist; that is to say, you alleged that stimuli and responses, which are mere physical energy inputs and effector events, would constitute your subject matter. Now I find you forced to employ a set of nonbehavioral concepts—namely, those of geometry. Furthermore, you do not employ them merely in the sense that you use your knowledge of geometry in designing the apparatus or what not. No, you employ them in a *substantive* way—that is to say, you use concepts from a nonbehavioral formal discipline as an essential part of the language with which you characterize the monkey's stimulus input. This seems to me to be contradictory to your expressed behaviorist aim."

I cannot imagine anyone voicing this complaint, and if anyone did, I cannot imagine any behaviorist taking the objection seriously. *Of course* he must at times employ mathematical formalism in other ways than as transformation rules in making theoretical derivations. Physical objects exemplify formal properties, and these properties are behaviorally relevant. It is just not possible to characterize certain stimulus inputs if one is precluded from employing the language of geometry. One way of viewing this is that we generally take the physical language, whether the ordinary physical-thing language or the theoretical language of physical science, as *including* certain portions of the languages of formal disciplines. That is, we do not forbid the physicist to write a Riemann integral, or the descriptive statistician to write down the expression for the gamma function, on the grounds that physics and descriptive statistics are supposed to deal with physical things which are "in the world," arguing that hence these sciences may not employ abstract or formal categories such as those found in mathematics or in an uninterpreted calculus. We are less accustomed to think of the formal features of a printed syllogism as a kind of "geometrical configuration," although Carnap made the point explicitly in his great work of 1934:

> Pure syntax is thus wholly analytic, and is nothing more than *combinatorial analysis,* or, in other words, the *geometry* of finite, discrete, serial structures of a particular kind. *Descriptive syntax* is related to pure syntax as physical geometry to pure mathematical geometry; it is concerned with the syntactical properties and relations of empirically given expressions (for example, with the sentences of a particular book). (p. 7)

> It is just as possible to construct sentences about the forms of linguistic expressions, and therefore about sentences, as it is to construct sentences about the geometrical forms of geometrical structures. In the first place, there are the analytic sentences of pure syntax, which can be applied to the forms and relations of form of linguistic expressions (analogous to the analytic sentences of arithmetical geometry,

> which can be applied to the relations of form of the abstract geometrical structures); and in the second place, the synthetic physical sentences of descriptive syntax, which are concerned with the forms of the linguistic expressions as physical structures (analogous to the synthetic empirical sentences of physical geometry, see §25). *Thus syntax is exactly formulable in the same way as geometry is*. (pp. 282-283)
>
> The sentences of syntax are in part sentences of arithmetic, and in part sentences of physics, and they are only called syntactical because they are concerned with linguistic constructions, or, more specifically, with their formal structure. Syntax, pure and descriptive, is nothing more than the mathematics and physics of language. (p. 284)

The point is made repeatedly and with beautiful clarity in several papers by Wilfrid Sellars,[3] although I shall content myself with only two brief quotations from his early (and insufficiently noticed) "Pure Pragmatics and Epistemology." Discussing the necessity of a pure pragmatics that avoids the philosopher's sin of psychologism, Sellars (1947a) writes:

> Today, then, the analytic philosopher establishes his right to attack psychologism with respect to a given concept if he is able to show that it is capable of treatment as a concept the nature and function of which is constituted by its role in rules definitive of a broader or narrower set of calculi. The issue was joined first over the concepts of formal logic and pure mathematics, and it can be said with confidence that the attack on factualistic and, in particular, psychological accounts of these concepts rest on solid ground. Logic and mathematics are not empirical sciences nor do they constitute branches of any empirical science. They are not inductive studies of symbol formation and transformation behavior. (And if, at a later stage in our argument, we shall find *formal* science dealing with language *facts,* it will not be because logic is discovered by a more subtle analysis to belong to empirical science after all, but rather because of a less naive analysis of the relation of language to fact.) This first battle was won because of the development of pure syntax. The concepts of formal logic and pure mathematics were clarified through being identified with concepts which occur in the formation and transformation rules definitive of calculi. These rules constitute a logic of implication and deducibility. In this stage of the battle against psychologism, an apparently clear-cut distinction arose between *symbol-behavior* and *formal system,* a distinction sometimes summed up as that between *inference as fact* and *deducibility as norm*. (pp.181-182)

And later in the same article he says:

> On the other hand, if we are asked, "Isn't it absurd to say that syntactical properties do not apply to symbol behavior?", we should find it extremely difficult not to agree. How, indeed, can we characterize an *inference,* for example, as valid, unless it makes sense to attribute syntactical properties to symbol-behavior in the world of fact? If we say that syntactical properties belong in the first instance to expressions in a calculus or language which is a model or norm for symbol behavior, can we then go on to say that in

[3] No philosopher or psychologist concerned with the "rules-and-facts" problems of semiotic can afford to leave Sellars's contributions unread or unstudied. See especially his "A semantical solution of the mind-body problem" (Sellars, 1953) and "Intentionality and the Mental [correspondence with Professor Roderick Chisholm]" (Chisolm & Sellars, 1958). See also Sellars (1947b, 1948, 1952, 1954, 1956, 1963).

the second instance they belong to language as *behavioral fact?* But to say this would be to put metalinguistic predicates into the object-language. Is there, then, no way out of our dilemma? Must we hold either that syntactical predicates are object-language predicates, or that syntactical predicates are not applicable to language as behavioral fact? Perhaps we can find a way out by drawing a distinction between language *as behavior* (that is, as the subject-matter of empirical psychology), and language behavior *to the extent that it conforms, and as conforming, to the criteria of language as norm*; or, in the terminology we shall adopt, between language behavior *qua* behavioral fact, and language-behavior *qua tokens* of language as type. (pp.184-185)

A difficult question which arises in connection with microanalyses of systems that perform logical and mathematical operations is the following: Suppose we deal with such a system, one which is clocklike rather than cloudlike, and we present a detailed causal analysis of the workings of the mechanism, including of course those *structural* and *configural* characteristics of the machine by virtue of which it "mirrors" or "embodies" logical and mathematical rules. If we do this microanalytic job adequately, it seems that we have performed the task of causal analysis, and yet we seem to have *left something out of our account,* namely, that which the mechanism is "achieving" or "doing." This puzzle arises in the philosophical analysis of conduct at least as far back as Plato (in the *Phaedo*) and continues to bother us today.

It seems not to be a mere matter of omitting adequate description of how the parts are arranged. It is obvious that one cannot be said to "describe" an ordinary desk calculator if he merely *lists* the parts, as such-and-so gears, levers, cams, cogwheels, and the like, even if he also gives a description of how a gear "works" (i.e., how it acts upon another gear in terms of the laws of mechanics) but omits to specify how the gears are physically arranged in the calculator. So, "calculating purpose" aside, it is clear that no one can claim to have provided a complete physical description of the machine if he leaves out an account of the arrangement of its parts, "how it is all put together." Let us suppose such a complete physical description to have been given. But let us suppose that the knower who carries out this "internal" analysis in terms of the principles of mechanics is a Martian visitor who uses a binary or duodecimal number system. (Or, even if he used ours, he might be unacquainted with the particular sign vehicles [= numerals] which we employ to designate the natural numbers.) That is to say, he has his own mathematical equipment (which is necessary for him to be able to solve the equations of mechanics involved in describing the inner workings of the machine); but he is not in possession of the rules of translation between his number system and ours, and therefore he might (conceivably) be forced to treat the Arabic numerals which are stamped on the keys, and which pop up in the register dials, as uninterpreted forms. If the calculator is structurally intact and functioning "properly," the rules of decimal arithmetic are perfectly embodied in the machine's structure, so that the operations of arithmetic are in perfect isomorphism with certain corresponding changes of state of the machine. The machine is—in the technical sense of the logician—a (physical) *model* of the interpreted calculus *arithmetic.* Thus, for example, punching the key marked T in the extreme right-hand column and then punching the key marked '+' is a physical operation sequence corresponding to the abstract specification "taking the successor of an integer." It is evident that the

Martian *could,* in principle, possess a mathematics adequate for a science of mechanics that would provide a "complete causal analysis" of the functioning of the machine, and *not* thereby (necessarily) understand the correspondence between the machine's structure (and structure-determined functional properties) on the one hand, and the Earthlings' numerical system on the other.

There is a sense in which, when the Martian has given his structural and functional analysis of the workings of the mechanism, he *has* "said everything that can be said," in the sense that nothing is "left out" of the causal analysis. But there is another sense, which is equally important, in which the Martian has *not* "said everything that can be said" about the properties of the machine, because he has not said that the machine "does arithmetic," or, less teleologically, that the machine's wheel movements constitute a model of a decimal arithmetic (= the wheel movements and positions satisfy the postulates of arithmetic). It is partly a matter of semantic convention how we choose to employ the locution 'saying everything that can be said.' But it is not purely conventional that one can distinguish between the following two kinds of *text* (I speak of 'text' because I want to emphasize that the following distinction is not a distinction of "mere descriptive pragmatics"):

1. A text is stated which consists of a conjunction of sentences exhaustively descriptive of the physical structure-dependent properties of the machine, and which suffice to entail all true statements about the machine's dispositions.

2. A conjunction of sentences (1) is stated, *together with all theorems which flow as consequences of the conjunction* (1) *given certain definitions.*

Now what *is* conventional or stipulative about the locution 'saying everything that can be said' is, of course, the possibility of stipulating that an individual who asserts the postulates also implicitly asserts the theorems. If anyone wishes to adopt this locution for certain purposes of logical analysis, I shall not complain of it. The fact remains that a text may contain the postulates without the theorems, or it may contain both the postulates and the theorems. And it is a familiar truth that while one in some sense "implicitly holds" the consequences of his postulates, in the sense that if he is consistent and rational he *ought* to believe the theorems that follow from them, the limitations of the finite intellect are such that we often do not hold all of the theorems which flow from our postulates because, for example, nobody has as yet succeeded in showing whether a certain well-formed formula is a theorem, or a counter-theorem, or even whether it is decidable. E.g., we do not know whether Fermat's Last Theorem is true or false; and we know that no one has as yet presented a valid proof of it, or a proof of its undecidability; so that it is somewhat misleading to say that we "believe it" or "hold it" or "know it to be true," supposing that Omniscient Jones knows it to be true, i.e., to be a consequence of the postulates of arithmetic. It is not, I think, an excessive reliance upon the usages of vulgar speech to ask that metalanguage stipulations avoid needlessly paradoxical consequences, such as that I am bound to hold that my late grandmother believed the number of her noses to be $-e^{\pi i}$

Now it might be said that whereas the Martian would lack (better, *could* lack) our semantics for interpreting the numerical sign vehicles that pop up in the dials

of the desk calculator, and if he were a particularly rigid or stupid Martian he might not develop insight into the translatability of the physical properties of the machine into a number system which was in turn transformable into his own, he could, nevertheless, "predict everything about the machine's behavior," because we have just assumed that he gives a complete mechanical-causal analysis of its micro-structural (and, as a consequence, micro-function) properties. And it seems evident that one who understands everything that happens in the causal order about any mechanism ought to be able to forecast—since we are assuming that the machine is completely clocklike and has no cloudlike "slippage" in its gears—all its dispositions. However, I believe there is an important sense in which even this is not quite true, unless stated very carefully and with all the necessary qualifications. The "results" of performing certain "operations" with the machine, definable in terms of what sorts of physical sign patterns pop up in the "answer" (cumulative bank) register are, after all, among the dispositional properties of the machine. And some of the strict uniformities (and statistical generalizations) in these "results" cannot, oddly enough, be predicted by a knower who has not made the cognitive identification of certain functional consequences of the machine's micro-structural features with the abstract concepts of arithmetic.

Consider the following example: A set of instructions is provided for carrying out division operations, and for recording their results, such that one obtains a sequence of "outcomes" (semantically uninterpreted by the Martian) that in fact constitute successive answers to the question "Is this integer prime?" We also have the Martian concurrently performing the task of keeping track of the proportion of such outcomes that have cumulated up to the *n*th integer, although of course he doesn't know that is what he is doing. Finally, we assume that the Martian knows logarithms (or at least that we can instruct him in the sheer mechanics of entering a logarithm table). Then it can be asked whether the proportion of outcomes of the "prime type" accumulated up to any point in this sequence of operations exceeds the reciprocal of the natural logarithm (i.e., the Martian is, so to speak, "empirically" examining Gauss's law of the density of distribution of primes). We now ask the Martian to predict, from his complete causal understanding of the machinery, how far along in the sequence he will have to go before he can be certain that the cumulated proportion of outcomes of the "prime" type will at some point have *exceeded* the Gauss approximation, instead of falling on the low side as it will at first. Now this number, Skewe's number, is

$$10^{10^{10^{34}}}$$

which is believed to be larger than the cardinal number of all of the atoms in the universe. So of course the Martian will never reach this value, before which—we don't know how much before—the prime proportion "flips over" so that Gauss's asymptotic formula errs on the high side of the actual value. We Earthlings, who *have* made the coordination—or for that matter an Earthling who *knows nothing about the machinery but only knows that the calculator* "*does arithmetic*"—can correctly state a lower bound for the number of such consecutive operations necessary to achieve this proportion of outcomes; whereas the Martian, or anyone else who has not made the coordination between the machine's structural pro-

perties and the axioms of arithmetic, could not make such a prediction. It seems obvious that this constitutes in some sense a genuine "cognitive edge," and that it is therefore false, or at least very misleading, to say that one who had described the internal mechanical structure, but has not made the explicit identification of the machine's states and operations with arithmetical concepts, would have "said everything that can be said." He would have said something which, given the appropriate explicit definitions and interpretations, *suffices to derive everything* that can be said, given the further assumption that he is an omniscient mathematician who is able to derive all the theorems that validly flow from the arithmetical postulates embodied in the structure of the machine. But if you don't say something that can be said, it is misleading to characterize your description as having said everything that can be said, even if what you have said is capable of entailing everything which can be said.

Nevertheless, recognizing this fact does not force us to postulate a "something more" going on *causally* in the machine, i.e., we do not infer from this that there is some sort of an additional arithmetic spook at work which sees to it that the calculator "obeys the postulates of arithmetic." Whether or not any such additional causal entity needs to be invoked depends upon whether our analysis of the situation amounts to a projection or a reduction, in Reichenbach's sense (1938, pp. 110-114). A desk calculator, an electronic computer, or—if physicalism be true—a human brain is a reductive complex of its elements. Nevertheless we have to insist that even in the case of a reductive complex, there is an important sense in which one may not have said everything that can be said about the complex, even though he may have said everything about the elements, and have included *certain ways of stating* everything about their relations, such that everything about the reductive complex follows of necessity from the statements which he *has* formulated. Thus if I recognize that a wall is a reductive complex of the bricks, and then I give the bricks numbers 1, 2, 3, and so forth, and state that Brick #1 is adjacent to Brick #2 and Brick #3 is located immediately above the first two and symmetrical with respect to them, and so forth, the vast conjunction of all true statements of this kind entails a "molar" statement about the wall. (They are not equivalent, since, as Reichenbach points out, the molar statement about the wall, while entailed by this conjunction of statements about the bricks, does not entail this conjunction; because the same molar statement about the wall is also a consequence of alternative conjunctions about the bricks.) What I have said about Bricks # 1, #2, #3, and so forth may entail the "molar" statement that the wall is 50 feet high; but if I do not make this latter statement, it is misleading to say, at least for certain purposes and in certain contexts, that I have "said everything that can be said." And it is at least theoretically possible for an individual to have a "complete understanding" of each of the statements about the elements and, depending upon the complexities of the structure, not to be (psychologically) able to derive a molar consequence that validly flows from these statements.

This is not an appropriate place, even if I had the technical competence, to enter upon a detailed consideration of the formal or structural relationships that must obtain between a physical system and a specified molar means-end process in order for the system to be capable of performing the specified process. And I

certainly do not mean to suggest that Professor Popper is unfamiliar with the problems in this area. I rather imagine he knows more about them than I do. Nevertheless, I must point out that his paper reads *as if* he believed a proposition which I am confident that he does not believe, to wit, "If a predicate 'P' designating a property *P* does not appear in a language adequate to describe a sequence of events related by causality, those events being considered at a certain level of analysis, it follows that the property *P* cannot, without inconsistency, be predicated of the system as a whole, or at another level of analysis." I do not myself know of any compelling reason for holding such a meta-proposition; and it is pretty clear that adopting it would generate some difficult (and, as I think, needless) puzzles. Example: Suppose we are talking political science, it would be a major lacuna in any characterization of Dwight Eisenhower to omit the statement that Eisenhower was a Republican. But even the most consistent identity theorist would consider it a category mistake to predicate of one of Mr. Eisenhower's cerebral neurons that the neuron was Republican. Must we say that since none of Eisenhower's neurons was Republican, therefore Eisenhower could not be such? Or, at the molar level, since none of Eisenhower's letter-forming actones (Murray, 1938, pp. 55-59, 96-101) (e.g., engraphing the mark 'a') may be meaningfully characterized as Republican, hence he wrote no Republican-oriented manuscripts? If the activities of the living human brain were—as I do not assert—completely "clocklike"; or if they were largely clocklike but with a certain irreducible element of "cloudiness"; or if they were extremely "cloudy," with only a small "clocklike" element present; in any case, no description of the cerebral processes at the micro-level, formulated in neurophysiological language, will include the predicate 'Republican.' It does not seem to me that this point about the appropriateness of certain predicates being confined to what one may loosely call "the whole person and his molar acts" has any relation to the question where the human nervous system is located on the cloud-to-clock continuum.

To stay away from the technical complexities of modern computer theory, consider an ordinary Hollerith punch-card machine. We have a batch of cards in a military personnel unit which are encoded in a certain way, i.e., the row and column positions have been, by some physical procedure, set into correspondence with properties, whether simple physical ones or extremely complicated social ones, of the military personnel whom the cards "represent." Professor Popper need not fear that I am surreptitiously avoiding the problem by shifting it backward to the encoding process. On the one hand, certain aspects of the encoding process do not involve "intentional mental acts" on the part of any encoder; but of course some do, and the ones that do will present a problem of microanalysis in relation to molar analysis of the same kind we are here considering. I am not, I trust, arguing circularly that there is no difference between the human brain and a Hollerith machine, a view I would vigorously repudiate. I am only saying that a Hollerith machine encodes information by a correspondence rule relating one set of properties, sometimes very complex ones, to another set, namely, a hole or non-hole at a specified locus on the card. That a human operator rationally intends the encoding process for his conscious purposes is true but irrelevant for my purpose at this point. Suppose that at induction a soldier fills in a

response-box set opposite to a named occupation on a checklist. We cannot presume that in filling in the box the soldier *has* to call up thoughts or images of his occupational activity, although of course he *might* do so. The more usual situation would be that a man who in civilian life has acquired the skills involved in making bread, and who has a pre-induction history of being paid to do this, will also be an individual who has acquired the verbal disposition to token 'baker' in response to the question 'What is your occupation?' Notice that there is no necessary overlap in the physical subsystems of the soldier's brain involved in these vocational activities and in his self-descriptive tokenings, nor does the self-descriptive tokening of 'baker' necessarily involve any concurrent or antecedent tokenings designating the activities of a baker. These dispositions are correlated in the English-speaking population by virtue of what R. B. Cattell calls "an environmental mould," that is to say, a cluster of topographically dissimilar dispositions which go together in a given culture or subculture by virtue of the fact that any human organism that learns the one will also have learned the other (Cattell, 1946, pp. 64-66, 74, 496; 1950, pp. 33-36). The correlation is similar to that which exists between a person's motoric skills in making an incision into living flesh and his disposition to respond verbally to the question "What is a Billroth II?" There is negligible overlap between these dispositions, either on their stimulus side or on their response side; and there is nothing about either one of them which suggests any appreciable overlap in the functioning of the cerebral machinery. The fact remains that they would be very highly correlated, because anyone who possesses certain incision-making skills at a given level of proficiency is certain to be a surgeon, and surgeons in the course of their training also learn to state verbally what a Billroth II is.

To return to our inductee-baker, he fills in a square box on an occupational checklist which the machine further encodes by punching a hole in a certain position on the card representing this soldier. Ditto for his height, weight, and eye color. (These can be coded mechanically, if desired.) Now suppose we want, for some strange reason, to select from a regiment all the enlisted men who are over six feet tall but weigh less than 180 pounds, who have blue eyes, and who were bakers in civilian life. The machine's board is wired accordingly, and the cards are run through the sorter ending up with a stack of cards in which are punched the serial numbers of all soldiers having this particular combination of properties. The functioning of this machine is very far toward the clocklike rather than the cloudlike end of Professor Popper's continuum. With a little care we can render it as clocklike as desired. Now suppose concerning any particular card which emerges from such a sorting process, we say, "Give me a detailed causal account of how this particular card happened to drop out during the sorting." This question can be answered in physical language describing the structures and processes of the machine without any reference to the vocational activities of the bakers, or to concepts of height, weight, and eye color. Nothing is left out of this causal account. If we start with the initial conditions on how the machine is wired, and how a batch of cards representing the entire regiment is punched, nothing *need* be said which cannot be completely expressed in terms of such concepts as brushes, electrical contacts, punched holes, the geometry of the coordinate positions at

which holes are punched, and the like.

Now this causal account of the card-sorting operation, which *leaves nothing out* (in one perfectly legitimate sense of the phrase 'leaves nothing out'), does not preclude a sentence of the following kind being literally true: "The sorter is picking out the cards of blue-eyed bakers over six feet tall and weighing less than 180 pounds." This is a perfectly good account of what the sorter is "doing." This sentence employs concepts which did not appear at the lower level of analysis of the machine's inner workings. Furthermore, it employs concepts which are *not translatable into the minimum vocabulary adequate to give an account of the machine's mechanical and electrical workings.* Is there any puzzle here? If so, it is resolved by recognizing the role of the previous encoding process, in which certain complex properties possessed by the soldier were set into a certain correspondence with loci punched on the cards. And corresponding to the fact that the cards can be repeatedly sorted is the logical particle 'and' joining the predicates 'blue-eyed,' 'baker,' 'weight less than 180 pounds,' and 'height over six feet.'

In order for me to be capable of making rational inferences or having intentions, it is necessary that my brain have a certain kind of structure. There are alternative physical arrangements that are equally capable of providing this necessary structure, the human brain being one of them. As long as the brain is capable of some kind of consistent encoding procedure, it can "represent" external facts, such as someone's being a baker, by nervous connections which, *when examined at their own level of analysis*, do not partake, however faintly, of "bakerhood." And even if the sequence of activation of individual nerve-cell dispositions were completely clocklike, this would not show, or even tend to show, that our beliefs, intentions, or volitions find no place in the world or that they have no causal efficacy.

It may be illuminating at this point to reexamine a famous puzzle about intentionality propounded by Sir Arthur Eddington (1929; the book is inaccessible to me at this writing so that while the basic puzzle is due to Eddington, the formulation of conditions is mine and may not accord precisely with his original setup). Suppose a man from Mars arrives on the earth mysteriously possessed of such a Utopian knowledge of Earthling neurophysiology that he is able, by a combination of behavioral and microtechniques (such as single-unit stimulation and the like), to give a complete causal account, *in neurophysiological terms,* of all the activities and dispositions of any given members of *Homo sapiens.* In particular, he observes (and was able to predict) that on November 11, 1918, great numbers of people in many cities of the world stand about in the public square waving their arms and shouting. Now, says Eddington, there is apparently "nothing left out" of this causal account; and yet the Martian would not know the most important thing there is to know about this social occurrence—namely, that these people are celebrating the armistice. This is true, if the Martian confines his attention to the momentary activities, but it is false if he allows himself to consider dispositions as well. There is surely no reason for saying, given Eddington's own assumption of a Utopian state of Martian neurophysiology, that the Martian is forbidden to include the dispositions of nerve cells in his description of the state of affairs. If these

micro-dispositions are included, I contend that Eddington is incorrect in saying that the Martian would not understand the "meaning" of the celebration. The reason is very simple: *Given a complete micro-account of the neural dispositions, one possesses all of the information necessary to construct a descriptive semantics*. He would, for example, know *that* (and he will also know *why*) persons waving their arms and shouting in the public square would be disposed to reply, if asked why they are carrying on in this crazy way, "The war is over." And, of course, his complete catalogue of neuronal dispositions would locate the word "war" in the descriptive semantic space, i.e., he would know what the word 'war' *means* to English-speaking human beings. Since we know that one can learn a language by recording the molar dispositions of its speakers (as, for example, an explorer or missionary *must* be able to do when he is the first visitor to a tribe of aborigines), a fortiori one would know the language if he knows all the micro-dispositions. Because, of course, the micro-dispositions entail the molar dispositions, but not conversely; and not all molar dispositions are realized in any finite behavior sample. Now it is perfectly true that the Martian is not *forced* to carry out any such descriptive-semantic research. He may, if he is only interested in neurophysiology, confine his explanations, predictions, and concepts to the micro-level. Whether such a confining to the micro-level "leaves something out" (in the causal sense) depends upon how far back in the causal chain it is desired to analyze the celebration. The immediate causal ancestor of a man's standing in Times Square and shouting would be, say, his looking at a newspaper headline "War is over." But the remoter causal ancestor is a complex of behavioral events at Compiègne, which the Martian would describe by a phrase in his language that is approximately synonymous with the English expression 'agreement to a cessation of hostilities.'

Can anything philosophically important about the mind-body problem, or the cloud-clock problem, be inferred from the fact that *if* physicalism is assumed true, the Martian *need not* pursue the causal chain that far backward but, on the other hand, that he *can* do so; or from the fact that he can predict "armistice behavior" successfully without tracing the causal chain back, confining himself to the momentary brain-cell dispositions; or from the fact that he could even infer the "meaning" of the celebration? I think not. These considerations do help to illuminate matters somewhat. Thus, for example, we are thereby reminded of the distinction between a statistical regularity of descriptive semantics (inferable by the Martian from nerve-cell dispositions) and the *non*psychological concept of a *semantic rule,* which is not a behavioral regularity but a prescription that the Martian formulates in his own Martian metalanguage. (See Sellars, publications cited in footnote 3 above.) A Utopian knowledge of the nerve-cell dispositions would be a Utopian knowledge of descriptive semantics, and a Utopian knowledge of descriptive semantics, together with a Utopian knowledge of the tokening dispositions of the celebrators, would obviously inform the Martian—*assuming he himself has the level of abstraction equipment which Eddington must presuppose in order for him to carry out the hypothesized microanalysis*—what the "content" of the celebration is all "about." All of which aids in dissipating the paradoxical flavor of the situation, but cannot help us to decide whether (a) physicalism and

(b) determinism are true doctrines about mind.

Professor Popper is disturbed by the notion that a "clockwork" view of the human mind implies that the behavior scientist would be able to "write" the symphonies of Beethoven through his knowledge of Beethoven's physical states, even though the scientist himself were completely ignorant of musical theory (Popper, 1966, p. 11). Why does this distress him? He says "all this is absurd." But is it really absurd? I will go him one better (partly to test the limits of my own convictions in the matter!). Consider the following example. Suppose a Utopian neurophysiologist studies the brain of a mathematician who is currently working on Fermat's Last Theorem. We will assume that this neurophysiologist knows the kind and amount of applied mathematics he needs to carry on ordinary calculations upon physiological measures, but that he is completely ignorant of pure mathematics, including number theory. Thus we assume that he has never even heard the phrase 'Fermat's Last Theorem,' let alone understands what it designates. Let us further suppose (with Professor Popper) that the cerebral mechanism has certain clocklike features but others that are cloudlike. In particular, let us suppose that there occur occasions on which the strengths of the neuronal activity in two systems of cell assemblies "competing" for command of the output channel are so close that a difference in only a few critically located "trigger" neurons firing or not will show up as a molar output difference. It is irrelevant for our present purposes whether at another level of analysis—say, by the physical chemist—this cloudlike feature arises from the quasi-random character of distributions of initial conditions of intra-neuron particles whose individual chemical and physical transitions are, nevertheless, completely clocklike; or whether it arises from a fundamentally indeterministic feature of nerve-cell action quantum-theoretical in nature, as has been postulated by some physicists and neurophysiologists.[4] In either case, the point is that a randomizing component is built into the functioning of our mathematician's cerebral system, super-imposed upon the clocklike features that are involved in his being thoroughly trained in mathematical manipulation (so that he always treats an exponent differently from a base, and the like). Suppose our Utopian neurophysiologist, ignorant of number theory, is able to show *at the micro-level* that there exists a set of alternative tokening dispositions, each of which is itself a chain of subdispositions to perform particular mathematical operations. That is, our neurophysiologist sees that the mathematician is "capable of" (= has non-zero probability of emitting) several alternative work-product sequences on a given day. Within each of these alternative chains, there are points at which the cerebral machinery functions clocklike,

---

[4] E.g., Bohr (1934); Eccles (1951, 1953, pp. 271-286); Eddington (1939, pp. 179-184; 1929, pp. 310-315; 1935, pp. 86-91); Jordan (1955, pp. 108-113); London (1952); Meehl (1958: "Determinism and related problems," chapter VIII, especially pp. 190-91; and footnotes 30, 31, pp. 213-215; and Appendix E, "Indeterminacy and teleological constraints," pp. 328-338; while I no longer hold the theological position presupposed in that discussion, the treatment of determinism and speculative brain processes still appears to me as essentially defensible; 1966, pp. 122-124); Pirenne & Marriott (1959); F. Ratliff (1962, especially pp. 442-445). For criticism of the notion that quantum-indeterminacy at the single-unit micro-level could be relevant to psychological determinism at the level of molar behavior or experience, see Schroedinger (1951, pp. 58-64); Grünbaum (1953); Stebbing (1937, pp. 141-242); and Popper (1966, section X, pp. 13-14).

and there are other points at which it functions cloudlike. (And even if it functioned clocklike at all points in the chain, the cloudlike selection of the *initial member* of a chain is unpredictable by the neurophysiologist.) So he doesn't know *which* of the chains will actually take place, but he can list all the physically possible alternatives. And if his psychophysiology is truly Utopian he can associate probabilities with each of these alternatives. (It is perhaps better to assume that the Utopian neurophysiologist is a considerably evolved species as respects his brain, studying a mathematician of *Homo sapiens.* Otherwise there may be information-theoretical difficulties involved in $Brain_1$ carrying out the requisite microanalysis of $Brain_2$ (Platt, 1966, pp. 147-149; the point has been made by several writers). These can presumably be avoided by setting no time limit on the neurophysiologist's derivation, so that he may continue work for months or years after the mathematician has quit. Or we may assume breakthroughs in computer engineering permitting superhuman computer brains. Or we may substitute "quasi-Omniscient Jones," who represses number theory, for the physiologist.)

Now, his microanalysis of each chain will obviously enable him to characterize the motor output—that is, the effector movements of the muscles of the mathematician's hand; and, consequently, from skeletal structure and biomechanics he knows what each virtual sequence of sign designs will be, i.e., *he knows what mathematical expressions the mathematician would write down,* if he carried out a given ("possible") cerebral sequence. Viewed thus, as the mere graphical residues of a molar class of finger movements (Neurath's "mounds of ink"), the potential work product might be devoid of meaning to our physiologist, yet its potential occurrence would be derivable by him from the Utopian microanalysis. So our Utopian neurophysiologist is able to list a set of mutually exclusive and exhaustive "behavior outcomes," namely, all the mathematician's potential work products for the day, although he does not know which one will actually take place but has only the probabilities associated with each. Finally, let us suppose that one of these "possible work products" is a valid proof of Fermat's Last Theorem. But, regrettably, it is a sequence having (for this particular mathematician) an extremely low probability; and it is not the sequence which in fact eventuates on the given day. ("The potential proof remains unactualized.") Having worked on the problem for several weeks, our unlucky mathematician becomes discouraged, and thereafter pursues other interests.

Now this mathematician was in some sense "capable of" a proof of Fermat's Last Theorem (assuming for the moment that a valid proof of this theorem does exist) but he in fact never discovers it. However, the neurophysiologist has now before him a list of alternative potential work products, only one of which ever came into being, and that actualized one is not a valid proof. The psychophysiologist takes the whole stack of hypothetical work products (each of which is directly derivable as a consequence of the effector movements terminating a chain of CNS events) to the Department of Mathematics. I remind you that the neurophysiologist doesn't know anything about number theory. He doesn't "understand what the mathematician is working on." Yet, the low-probability valid proof, which was never actually carried out by the mathematician, would be

recognized as a valid proof by the Department of Mathematics. Thus the neurophysiologist in some sense could "discover" a valid proof of Fermat's Last Theorem without understanding mathematics, by studying the brain of a mathematician who, while in some sense potentially capable of developing such a proof, never in fact does so. I readily agree that this sounds counterintuitive. But I do not see anything contradictory about it. And I think the reader will agree with me that it is interesting.

I have argued above that statements about human behavior or experience which attribute causal efficacy to reasons have a meaning which should be acceptable both to a philosopher-logician and to a determinist psychologist, but that such statements are elliptical so that unless carefully unpacked they are likely to be misleading. Thus I have said that, strictly speaking, a valid argument, considered as a certain formal structure (an abstract universal) is not an event "in the world," at least in any ordinary sense; and therefore it cannot function as a causal agent with respect to an event, e.g., a human locomotion, manipulation, or phonation. The ontology of universals, the reality of abstract entities, and the more technical aspects of the traditional nominalism-realism debate are not—it is hoped—relevant, because a serious discussion of them is, regrettably, beyond my competence. My colleague Professor Maxwell thinks I am wrong, or at least terminologically ill-advised, to say that logical relations (such as deducibility) are not "in the world." As he—rather compellingly—puts it, *"Everything* has got to be 'in the world'; where else *could* it be?" Professor Popper (1968) even writes recently of a "third world," whose denizens are abstract ideas. I confess I do not understand Sir Karl here; but perhaps Professor Maxwell's demand is met by my agreement with Carnap and Sellars on linguistic structures [see text pp. 21-23 *supra* with references to Carnap, 1934; Sellars, 1947, and footnote 3]. In any event, by saying that there are Platonic universals but they are not in the world, I have made my position more difficult, and Professor Popper's easier, to maintain. I have argued that when a bit of rational behavior is being fitted into the causal framework, the question whether certain logical categories (such as the category Illicit Distribution) are required in formulating the behavioral laws or quasi-laws depends upon the level at which the behavior analysis is being conducted. If we are attempting to formulate psychological laws either in mentalistic language or in molar behaviorese we will find such formal categories indispensable, because we will be unable to characterize a stimulus class which functions as a discriminative stimulus for such verbal responses as 'Illicit Major' on the part of a logician-subject *unless that stimulus class is characterized by reference to its syntactical structure.* Whether the mentalistic or molar-behavioristic psychologist actually employs the logician's *terminology* or not is irrelevant, inasmuch as he will be driven, in his account of the subject's behavior or experience, to introduce a specification of the syntactical features of the stimulus input, which specification will in fact *be* the logician's definition of 'Illicit Major.' But I have also maintained that the same is not necessarily the case, given a complete Utopian micro-description, although the complete Utopian micro-description will *entail* (within the nomological network) the same syntactical statements at the molar level which would have to be invoked by the molar behaviorist in predicting or

explaining rational behavior. This is because the whole organism and its molar activities are reductive complexes of the micro-structures and micro-events, and hence the statements about the whole person follow from the statements about his component parts and part processes, analogously to the way in which statements about a wall follow from conjunctions of statements about the bricks. But we have also seen that there is an important sense in which, unless one *asserts* these molar statements which are consequences of the statements about the elements, he has not literally "said everything that can be said" about such a reductive complex.

The anti-determinist or, perhaps more strongly, the ontological dualist may object to this analysis with the following: "You say that your refurbished behaviorism, including as it does a physical$_2$ microanalysis, and a recognition of the molar-indispensability of certain logical categories such as *valid form,* does justice to the logician's legitimate claims, while still maintaining physicalism and determinism in the domain of mental life. In this you attempt to please both parties; you want to have your cake and eat it too. I do not know whether the hard-nosed behaviorist will buy this, but I, as a firm believer in the genuine efficacy of reasons, cannot buy it. Because, while you tell me that the micro-account *entails* those statements at the molar level which are characterizations of stimulus inputs as logical forms, the fact remains that you also maintain the dispensability of concepts like *valid reason* in a complete causal analysis. Because while you admit that one who fails to assert some of the consequences of those statements which he does assert has failed (in a sense) 'to say everything that can be said,' nevertheless it remains true that you hold it possible to present a complete causal account of human actions without reasons entering the causal chain *as reasons.* That is, you maintain that it would in principle be possible, within a Utopian neurophysiology, to detail the processes in a person's brain confining oneself to physiological descriptions at the level of 'neuron-language,' such that the resulting molar output, e.g., punching somebody in the nose, could be predicted and completely understood causally at this level of analysis; and it is obvious that no reference to the good reasons he may have had for punching somebody in the nose would occur in such a micro-causal analysis. This is what I mean by insisting that you are depriving rules and reasons and validity of all genuine efficacy in human affairs. If you can give a complete causal account of what a person does and why he does it without at any point mentioning the reasons *for which* he does it, then it seems to me that you have, in effect, eliminated the reasons from any significant role. You throw a sop to me and my friends the indeterminists, emergentists, Cartesians, etc., by telling us that certain conjunctions of micro-statements entail molar-level statements which—given suitable metalanguage definitions of logical notions like 'implies' and 'negates'—in turn entail statements about a person's *reasons,* in our full sense of 'reasons.' But you also insist that one *need not do this* in giving the complete micro-causal account. You, so to speak, 'permit' us to mention reasons; but you insist that you yourself are not *compelled* to mention them. But, surely, if they need not be mentioned, they are dispensable. And this we cannot admit."

Since I myself admit—nay, I insist, as against a certain kind of behaviorist—that reasons are psychologically efficacious, i.e., that the hearing of reasons and

the thinking of reasons and the tokening of valid arguments play a role, and for rational men may play the crucial or determinative role, in the guiding of their actions and utterances, it is the more obligatory upon me to answer this objection. It seems to me that the core issue here can, without prejudice, be formulated thus: Has one "dispensed with" the causal efficacy of a configural property of a physical state or system as playing a role in a causal explanation *whenever he avoids explicitly characterizing that configural property*, confining himself to description at a lower level of analysis ("lower" in the sense of a reductive complex), provided that (a) what he *does* assert in his lower level description can be shown to entail nomologically the configural statements and (b) if asked, he concedes—as he must in consistency—that these configural consequences are entailed by his lower level statements? Is there not a considerable element of conventional or stipulative usage involved here, about which it is pointless and fruitless to argue? One who describes a physical system omitting dispositional statements about it which flow as necessary consequences of the statements he has made, might be said, on one convention, to have "left something out," because he did not say everything that could, and (strictly speaking) everything that *must,* if the question is raised, be said. But so long as he is not inconsistent, so long as he is quite willing to *admit* the necessary consequences of what he has said, and those consequences of course *include* the entailed configural properties of the system, the locution "He thinks these configural properties are irrelevant to an adequate account" is surely misleading. In what sense can I be said to think that any feature of a physical system is "irrelevant," if I concede that this feature is a necessary consequence of features to which I have attributed relevance, and further, that if the system *lacked* these (entailed) configural features, its "output" characteristics (e.g., tokening an implied conclusion when one has tokened the premises) would be different from what they in fact are?

In assessing the conventional element in whether we would think it convenient and clarifying to say that such a scientist "leaves something out" or "considers something causally inefficacious," one consideration might be whether the physical system includes a subsystem which functions as a kind of controller, guider, evaluator, or selector, with respect to another subsystem, such that, among the intermediate or molar-level theorems that flow from the axioms of which the system is a model, there is a statement which says, roughly speaking, that the monitor or selector system will "accept" or "reject" a certain product or message from the monitored or controlled subsystem, depending upon whether that product or message possesses or fails to possess such-and-such formal properties. Thus, for example, when we program an electronic computer to perform certain computational checks upon its own work and to report to us the presence of inconsistencies; or when we program it to inform us that our own program is itself defectively written—in such cases it seems very natural, and not just a computer engineer's whimsy, to use the connective 'because' in sentences like the following: "The computer rejects these data *because* they include entries in a correlation matrix which exceed unity." It is true that in this kind of case it is also possible to describe and explain the operation of the monitoring or evaluating subsystem in micro-terms. But it remains true that we can identify two such functional sub-

systems, and we can correctly (and literally) say that the monitoring subsystem *classifies* the states and outputs of the monitored subsystem with regard to their possession or nonpossession of certain formal properties. It seems to me that one can arrange a continuously graded series of physical systems, each of which is a physical embodiment of certain formal rules (i.e., each of which is, in the logician's sense of the word, a "model" or provides an "admissible interpretation" of a formal calculus) from one extreme at which it would be a very marked departure from ordinary usage to employ the connective 'because' followed by a characterization in terms of validity or logical structure, to another extreme at which a failure to include this intermediate or molar-level characterization would be looked upon as some sort of prejudice or inadvertence. Take, for instance, the case of a beam balance, which we do not ordinarily think of as performing logical or mathematical operations. We place three one-gram weights in one pan, and we place two one-gram weights in the other pan, and we observe as a causal consequence of these physical operations that the beam becomes and remains nonhorizontal. It would not ordinarily occur to anyone to describe this state of affairs by saying, "The balance tips *because of a truth of arithmetic,* namely that $3 > 2$." But there is a perfectly legitimate sense in which such a statement would be literally correct. If there is a Platonic sense in which the truths of arithmetic are not "in the world," there is another sense in which they are, namely, that since these theorems are analytic, all physical objects do in fact exemplify them. (Cf. Wittgenstein, "It used to be said that God could create everything, except what was contrary to the laws of logic. The truth is, we could not say of an 'unlogical' world how it would look. To present in language anything which 'contradicts logic' is as impossible as in geometry to present by its coordinates a figure which contradicts the laws of space; or to give the coordinates of a point which does not exist. We could present spatially an atomic fact which contradicted the laws of physics, but not one which contradicted the laws of geometry." [1922, Propositions 3.031-3.0321]. See also Popper, 1962, pp. 201-214.) If we move along this continuum of "rule representtation" from the beam balance (which "exemplifies," "instantiates," "physically embodies" the axioms of arithmetic, as well as the formalism expressing the laws of mechanics—the former necessarily, the latter contingently) to the ordinary desk calculator, it still seems somewhat inappropriate, but much less so, to characterize its operations by using the connective 'because' followed by an arithmetical truth. I note that even here we are more likely to do so in an extreme or "special" instance such as an inadvertent division by zero, where we say, "The machine keeps running and doesn't pop up with an answer, because you divided by zero." (We here correlate the mechanical fact that it would "run forever" if the gears didn't wear out with the arithmetical notion that

$$\frac{N}{X} \to \infty \text{ as } x \to 0$$

or roughly put, that if division by zero were allowed the answer would be "infinity.") The fact that the gear wheels in the calculator are toothed in isomorphism with the decimal system, and that they are arranged from left to right in isomorphism with the way in which we place numerals in the decimal system to

represent the powers of 10, facilitates our intuitive appreciation of a more explicit embodiment of the "rules of arithmetic" in the machine's structure than we readily feel in the beam balance case. Just how natural it seems to employ the locution 'because' followed by an arithmetical truth seems, in the case of a desk calculator, to depend partly upon the complexity of the operation involved, a psychological aspect which does not reflect any fundamental physical or logical difference. For example, suppose I am given a printed instruction the rationale of which I do not understand, as follows: I first set a number into the upper (cumulator) register; then I proceed to subtract the consecutive odd numbers, 1, 3, 5, … until I get all zeros in the register; then I record the number which appears in the lower (counting) register. If I now clear the machine and operate upon this recorded result with itself through the multiplication key, the machine presents an "answer" in the upper (cumulator) register, and that answer is the number that I started out with. Suppose I am baffled by this, and I ask the question "What is the explanation of this remarkable mechanical phenomenon?" I would probably be satisfied, unless I were specifically interested qua mechanician in the internal workings of a desk calculator, by someone's saying, "Oh, that happens because it is a theorem of arithmetic that the sum of the first *k* odd numbers is equal to the square of *k*." It seems to me that whether one views this use of 'because' as literal or figurative is a matter of adopting a semantic convention, rather than a psychological, ontological, or epistemological issue about which there can be a genuine cognitive disagreement. If, for example, one were to require, in stipulating what constitutes a legitimate use of the word 'because' followed by a theorem of some formal science (logic, set theory, arithmetic, differential equations) that the physical system should in some suitable sense be *tokening the theorem* (rather than merely exemplifying it), then he would say that the use of 'because' in the present instance would be incorrect usage (or, at best, metaphorical). But there seems to be no compelling reason for adopting such a stringent stipulation regarding the word 'because.' We employ logic and mathematics to describe the world, whether in its inanimate or animate features. A configural feature of a physical system, whether animate or inanimate, is literally characteristic of it. A formal theorem, whether of logic or mathematics or set theory or whatever, that is exemplified by the system's states and lawful transitions is, I submit, literally attributable to factual (contingent) structure-cum-events of the physical order. Unless some strong counter-consideration were advanced, such as the danger of confusion or of anthropomorphic projection (e.g., of feeling states of experienced motives) into an inanimate system, it is hard to see why such locutions should be conventionally forbidden.

If it is now objected that 'because' cannot be stipulated as allowable usage without doing great violence to both ordinary language and technical conventions, on the ground that the rule exemplified is not in the causal chain, I am at a loss how to reply beyond repeating what I have already said, to wit, that while the *rule* is not in the physical order, an *embodiment* (model, satisfier) of the rule *is* in the physical order. I would say further that the legitimate element of "necessity" which most logicians today would be willing to concede (in spite of Hume) is clearly present in this type of situation. That is to say, if we reconstruct a post-

Humean notion of causal necessity as a combination of (a) logical necessity, (b) analysis of reductive complexes, and (c) the distinction between fundamental nomologicals and derivative nomologicals, then we properly assert that the calculator gives the answers which are *arithmetically necessary*, and that it does so "necessarily," given the presupposition that the laws of physics hold and that the machine is not broken, worn out, or the like.

When we move to the modern electronic computer, an additional element enters which makes it still more natural to refer to logical and mathematical theorems in explaining the machine's behavior, namely, that the machine contains a physically identifiable subsystem which stores "instructions" of a nature less generic than the ever-binding laws of logic and arithmetic. And as these instructtions become more and more complicated, as when we instruct a computer to examine a certain result with respect to some property (such as whether it is odd or even, or whether it is greater than a certain value) and, depending upon the result of this examination, to operate upon this result in one or another way, then we feel quite at home with such explanations phrased in terms of rules of logic and arithmetic and the word 'because.'

There is, however, still something lacking in the computer which we might wish to require before we employ the word 'because' followed by a reference to a logical rule, namely, a physical subsystem which corresponds to a psychological *motive.* If we can distinguish motives from nonmotivated intentional states, one can say that a computer has intentional states, i.e., applies rules, but does not have motives, i.e., it does not desire things. I myself can discern no division point (other than the phenomenal or consciousness criterion) on the complexity-of-goal-seeking dimension which is other than arbitrary. When Samuels's (1959) checker-playing computer (which *learned* to defeat him, the programmer!) examines a set of possible eight-move sequences, and selects the initial move of that sequence which optimizes certain features of the resulting position, I would insist that this is an unquestionably intentional, goal-directed, criterion-applying process, *except* for the "sentience" component. The same can be said for the Logic Theorist computer program, which cooked up a shorter and more elegant proof of the *Principia Mathematica* Theorem *2.85, done in three steps rather than Russell and Whitehead's nine, relying on fewer axioms, and rendering a certain lemma superfluous (Newell, Shaw, & Simon, 1957, 1959; and see generally: Hunt, 1968; Newell, Shaw, & Simon, 1958). (This kind of thing weakens Professor Popper's argument from "novelty" or "creativity," I think—even though the example deals with the propositional calculus where neither Church's Theorem nor Gödel's troubles us.)

Here again one has a problem of stipulation. I myself would include, as a necessary ingredient of "desire," the subjective phenomenal, experiential component, which I presume to be lacking in an electronic computer (regardless of the logical complexity of its "intentional" features). As to "goal," I am quite neutral. As to "selection," I say the computer *selects.* At this point the sapience aspect of the mind-body problem borders on the sentience aspect, which is not the subject matter of this paper. It is, I think, arguable that adopting a convention requiring a raw-feel experiential aspect as a necessary component of the construct *motive*

would be very inconvenient. It might preclude the animal psychologist or ethologist from attributing motives to animals at certain levels, where the phylogenetic continuity in many goal-directed aspects of behavior does not seem to be interrupted, but where it becomes increasingly dubious whether any such subjective or raw-feel component is present. But more important, while one may entertain (as I do) grave doubts about the validity of considerable portions of the Freudian picture of the mind, I am prepared to argue that one of its core characteristics, the same basic idea of the controlling influence of motive-like variables or states which are not reportable by the subject as having an inner-phenomenological aspect, seems rather well corroborated. It would be very inconvenient, both theoretically and in clinical practice, if we were forbidden to refer to an individual's motives except in those cases in which he is able to give an introspective report of them. I suspect that most psychologists would, like myself, put greater emphasis here upon theoretical generality than upon vulgar speech. Hence the preference to use 'motive' after the manner of Freud or Tolman, the conscious/unconscious distinction being made not by noun choice but by an adjective (conscious motives are contrasted with unconscious motives). But of course if someone wants a noun (e.g., 'desire') that always means *conscious motive,* well and good. These semantic preferences are stipulative, and it is silly to hassle over them. What is *not* stipulative is the empirical finding that much of human behavior is controlled by internal state variables or events that (a) have most of the usual causal properties of reportable motives but yet (b) are not reportable as having a subjective-experience aspect. It is not easy to improve on Freud (1915/1957) and Tolman (1932) in spelling this out.

If a human brain, like an electronic computer, has a structure which makes it susceptible of storing instructions concerning the allowability of certain kinds of transitions, it would seem appropriate to say that a person accepts a particular argument *"because* it is valid" and rejects another one *"because* it is invalid." That one could carry out a complete causal analysis without referring to these logical meta-categories (because he might do it instead at the micro-level) does not invalidate the literal truth of this statement, although it is commonly thought to do so. One way of seeing this is to put the question "Would the argument be accepted by the individual if it were not formally valid?" The answer, of course, is that if the argument were not formally valid, then it follows (from the fact that the micro-laws entail the macro-laws and the macro-state is a reductive complex of *the micro-states that—literally—compose it*) that the cerebral mechanism would reject the argument. Hence, if the critic says, "According to your view the validity of the input argument *makes no difference in what happens*," we would have to reply that the critic is simply mistaken in saying this. Because it can be shown from the microanalysis that if the input argument were invalid, the macro-behavior, i.e., the logician's tokening response, would be to say, "This argument is an Illicit Major." Any configural property of an input which "makes a difference," in the sense that if it were lacking, the individual would say "No," but if it is present the individual will say "Yes," surely must be said to "make a difference" in the most stringent use of that expression.

While the complexities of the reconstruction vary widely, and while the

presence of a state or event which constitutes a "guiding motive" (*properly* so-called) makes a great difference, the fundamental point I am making seems to be exemplified both in inanimate and animate contexts; and, within animate systems, is exemplified both in the "psychological" and the "purely physiological" domains. In physics there are problems such that the state to which a system moves can be derived by alternative methods which are not contradictory but which do represent analyses at different levels of description. Thus, we may invoke highly general principles of a quasi-teleological sort (such as conservation principles or least-action principles), but we may sometimes achieve the same result (less easily and elegantly) without invoking these principles, by proceeding at the micro-level of causal analysis. Or, again, a certain theoretical concept may be one which has a summary function, or which characterizes a complex configuration by reference to certain summarizing quantities, so that the attribution of a certain value of the summarizing quantity follows necessarily from what would be a huge conjunction (or disjunction of conjunctions) of statements about the components. In such situations, there is a noncontroversial sense in which one who omits mention of the summarizing quantity may be said to have "left something out of his description," because he did not say everything that might correctly be said. Putting it more strongly, he omitted saying something that is necessarily true on the basis of those things he has in fact said. But I think we should not say that he has given a *defective* causal account because of this omission, inasmuch as the omitted statement concerning the summarizing quantity is virtually present (in the sense of logical entailment) given the nomologicals, and/or explicit definitions, in the statements he has made. Example: I place a block of ice in the center of a room which is being kept warm by a roaring fire in the fireplace. An omniscient physicist provides me with the monstrous conjunction of micro-statements regarding the collisions of individual air molecules with the molecules at the icecake's surface and a blow-by-blow quantum-mechanical account of the manner in which the intra-molecular forces holding each particular molecule in its position in the ice crystal are counteracted, so that this molecule becomes free of the crystal, i.e., becomes part of the fluid. At no point does he employ the terms 'crystal,' 'melt,' 'fluid,' and even his references to the internal geometry of the ice crystal are clumsily formulated by a complicated conjunction that avoids reference to planes, lattices, and the like. Now the true statements involving all of these macro-and intermediate-level concepts are, given the explicit and contextual definitions of these terms in physical theory, to be found among the sentences that follow nomologically from this vast conjunction of sentences characterizing the micro-states and micro-events which he does utter. I do not believe that anyone could object to such an account, other than on aesthetic grounds or because of its cumbersomeness. That is, one could not object by saying that the causal account has been rendered somehow incomplete by the failure of the physicist to include mention of all the sentences, and therefore the words that would normally occur in such sentences, at a more molar level of causal understanding. No one would object to this account by saying, "That's all very well, but it won't do as a complete causal explanation of what took place, because you have described the situation as if the difference in temperature

between the cake of ice and its surroundings was irrelevant, i.e., that such a temperature differential had no efficacy, that the fact that the ice was colder than the fire-heated air *made no difference*." To say that this micro-account is defective because it suggests that the temperature difference between the ice and the surrounding air "had no effect," "was irrelevant," "made no difference," or that it was some kind of a "supernumerary," "mere parallel," or "epiphenomenon," would be misleading. The complete characterization of the situation concerning the surrounding air, and the causal explanation of that situation by reference to the chemical changes taking place in the fireplace, obviously do "make a difference," in the literal sense that if those circumstances had been other than they in fact were, the micro-events described in the huge conjunction of statements about the freeing of the individual molecules of water from their crystalline state would not have been true. The conjunction of a vast set of sentences about the molecular motions of the air together with the conjunction of statements about the molecules in the ice, entail a statement that the summarizing quantity known as "temperature" will be higher in the one than in the other.

Consider a nonpsychological case in the animate domain. A biochemist describes the processes which go on in a man's blood chemistry over a period of years at the biochemical level. A physical chemist explains the micro-details of how these various values of blood concentrations bring about a deposition of lipids in the intima of the coronary artery. A physicist gives us a detailed micro-account (in terms of Euler's equations, etc.) of the hydrodynamic situation at this site, resulting from the narrowing of the arterial lumen. A physiologist provides an explanation of what happens in the individual cells of the cardiac muscle itself as a consequence of the reduced oxygen supply and deficient rate of removal of metabolic products, including a micro-characterization of processes which a cytologist, *if asked,* would recognize as constituting "death of the individual cell." Now one might legitimately complain of this account (especially if he were in the life-insurance business, or a close friend of the family's, or the attending physician) that it fails to state something that was literally true, something that should appropriately be said at another level of analysis, namely, that the patient suffered a myocardial infarction as a result of a coronary occlusion. But it would be misleading for this objector to say that the team (physical chemist + biochemist + physicist + physiologist + cytologist) had "left something out of the account," if by that is meant that there was some further event, entity, or process at work influencing the chain of causality and that this something had been omitted from the description. It would be very wrong of the critic to say, "You have described this as if it made no difference in causing the man's death that he had a coronary occlusion which produced a myocardial infarction." Nothing of the sort. The conjunction of lower level statements made by our five-man basic science team, when taken together with the explicit definitions of the science of pathology, *is* an assertion that the patient suffered a coronary occlusion leading to a myocardial infarction.

Again, suppose we consider two flywheels having the same mass but different physical dimensions, whereby one has a much larger radius of gyration than the other. We inquire about the torque necessary to accelerate these flywheels to a

specified rate of revolution, but we do this by considering the transmission of the applied force through the material particle by particle, and by literally summing these billions of components, rather than performing the usual integration analytically from the geometry of the two flywheels. Of course it turns out that in spite of their equal masses, a greater torque is required to achieve a stated angular acceleration in the case of the flywheel whose mass can be considered as concentrated at a greater distance from the center. But we do not explicitly mention this in our causal explanation. Would anyone object to our causal account, saying it was "incomplete" because it "treated the radius of gyration as something lacking in relevance or effect"?

Whether or not one would "normally" or "naturally" employ locutions of an intentional or quasi-purposive type in vulgar speech is, of course, largely lacking in scientific or philosophical interest. So long as we understand the causal and logical categories and their relationships to one another in the various contexts, whether we opt for one or another label is uninterestingly stipulative. My intuittions about what the usage of vulgar speech would be in a given setting are, like everyone else's, armchair speculations based on anecdotal impressions and lacking in such scientific support as might be obtained through a properly designed sampling procedure with the best available methods of psycholinguistic investigation. What sentences the alleged "plain man" would be willing (or reluctant) to token in regard to the behavior and "mental processes" of a digital computer which beats him at a game of checkers is one of the dullest topics imaginable, and cuts no philosophical or scientific ice so far as I am concerned. My own references in this paper to what (in my armchair opinion) we would "ordinarily" or "naturally" say are intended pedagogically and psychotherapeutically, and nothing really hinges upon these educated guesses of mine. I am not interested in armchair psycholinguistics, whether Oxbridge or Minneapolis style; and I should not dream of deciding a scientific or philosophical question on the basis of what I guessed would be the verbal behavior of a (hypothetical) uninformed layman were he asked to think about difficult and obscure matters which he does not in fact think about, and lacks adequate conceptual equipment to think about.

It is, nevertheless, instructive from the standpoint of curing one's own intuittive resistances to pinpoint their source, considering a variety of examples with an eye to analyzing carefully those in which one experiences considerable ambivalence with respect to the use of quasi-mentalistic or quasi-purposive labels as applied to an inanimate system. Speaking for myself (and I invite the reader to consider whether this may be true of him also), it is my impression that when the element of sentience either is excludable in high probability or is irrelevant (because it is not supposed to be causally efficacious, or because it is clearly present in *both* systems under comparison), the human/subhuman or even animate/inanimate dichotomy sometimes receives less subjective weight in our readiness to employ purposive or intentional locutions than does the question whether the process under study has in it some features of "matching" or "comparison" of the actual with the ideal. So that when an inanimate physical system has a structure which enables it to function as a kind of "judge" or "comparator," which "applies criteria," we readily employ mentalistic verbs such as 'sort,'

'classify,' or 'test'; whereas we are reluctant to employ such language, *even with respect to an organic system,* if this element of judging, of comparing, of determining whether something satisfies a condition is utterly lacking.

It is true that in such cases we are aware of the fact that a human mind constructed the inanimate mechanism for a certain purpose, and it is sometimes argued that *this* understood origin and (human) purpose is what justifies the use of such mentalistic language in speaking of a calculating machine. There is no doubt considerable truth to this, and I have no wish to play it down. However, I wish to maintain that such an empirically based comprehension of the human designer's purpose in building the inorganic machine is, while typically present, not a necessary condition for properly applying some (not all) of these "criteria"-flavored words. For example, there are machines the function of which, *part by part,* a layman with a rudimentary understanding of mechanics and mathematics could be brought to understand, but the overarching industrial or scientific purpose to which they were put might baffle him. He might not have anything like an adequate comprehension of the desired "end product" envisaged by the engineer who built the machine to satisfy certain human motives, but he might nevertheless be capable of *characterizing the properties* of the end product satisfactorily. Or, to take a science-fiction example, suppose we find, on geological excavation in the pre-Cambrian rock strata, several complicated apparatuses on which no written instructions are provided in a language we understand, but there is a small plate showing a picture of what appears to be a large bird fastened in the machine. We get ourselves an adult male ostrich and find that he "fits into the machine" with a little adjusting here and there, and when we start the thing running it turns out that what it does is to pluck prime-numbered feathers in a line running down the ostrich's back. This would be pretty spooky, and we might have a very difficult time understanding what a pre-Cambrian somebody was up to, but after experimenting with several ostriches on several of these "extinct machines," we would be entitled to say that, odd though it is, there is good evidence that the "purpose" or "function" of the machine was to pluck out the prime-numbered feathers from ostriches. Point: One does not need to understand the overarching "why," the *ultimate* "end in view," of the maker of the machine in order to infer something about the characteristic end product, as *proximate* "end in view," which results from the machine's operation. In fact, we do not always presume a designer (or plan or prevision) when asking quasi-teleological questions; witness the atheist zoologist or anatomist who undertakes research to answer the question "What is this organ *for*?"—a question which is, given careful formulation, surely sensible aside from one's views on natural theology. The most incisive and illuminating discussion of this problem that I know of is by Nagel (1961, chapter 12; see also Hempel, 1959, reprinted with alterations in Hempel, 1965).

At the risk of boring the reader, let me consider one last example, to highlight the problem of causal analysis in relation to abstract universals, when the latter are allowed to go unmentioned in a (purportedly complete) causal account. We have a simple industrial testing machine, a plate with an elliptical-shaped hole in it, the hole being made slightly larger than any of a batch of elliptical-shaped tiles which the machine is to "test." The machine has "arms" that place each tile in

position above the hole, and then proceed to rotate the tile slightly in both directions from its initial position, the rotation being smooth (or at least by steps of very small angles) so that an optimal placement will not be inadvertently "missed." The relationship between the distribution of sizes of the ellipses and the amount of "play" in the hole is such that an elliptical tile whose major and minor axes are in any ratio between 6:8 and 7:8 will be capable of falling through the hole and dropping into a collecting box below. Otherwise the machine throws the tile aside. Now suppose someone provides a detailed account of exactly what happens in the sequence of operations involved in testing 100 consecutive tiles on this machine. He describes the form of the testing slot by stating the coordinates of a very large number of points on its edge (located, say, 1 millimeter apart). Thus he does not write any mathematical function in the familiar algebraic form, although he does in effect write a function for it in the sense of providing a finite set of ordered pairs of numbers. After having thus "tested" 100 tiles, ending up with 97 of them in the box (these having passed through the test slot successfully) and 3 "rejects," and having given a detailed account of the sequence of events as each tile was being tested, our nonmathematical mechanic offers this as a complete causal account of what happened. But now a critic advances the following: "You are assuming, in your alleged complete causal account, that the elliptical shape is of no relevance, that the relation of the major to minor axes in the tiles makes no difference." He makes this objection on the ground that we have not made any reference to the word 'ellipse,' or said anything about major and minor axes. Would this criticism be valid? We would admit that critic is now pointing out something which is literally and physically true about the sequence of events, namely, that *whether or not a particular tile ends up in the box or as a reject depends upon whether it meets or fails to meet a certain geometrical specification.* This specification, that of 'being an ellipse' (with a certain range of tolerance permitted in the ratio of the axes), has been "left out" of our account. So we have not said everything that could be said. But would one infer from this that mechanism or physical determinism must be false as an ontology of testing machines, or that a reified platonic universal (some sort of ideal elliptical tile laid up in heaven) must get into the act somewhere to see to it that things go properly?

An ellipse is an abstraction, a universal belonging to the domain of a formal science, which certain material objects may "model" (in the usual sense that a model is a set of entities that satisfies a calculus). What we have presented, in our allegedly complete physicalistic description of the slot, when we specified the coordinates of the points running along the edge 1 millimeter apart, is a set of statements which collectively entail that a tile which passes through is of elliptical shape, with axes in the range 6:8 to 7:8.

Keeping in mind my exclusion of the two considerations beyond the scope of this paper (the subjective-experiential aspect of the mental and the motivational or purposive), let us imagine a critic who adopts Popper's view about the causal efficacy of universals to advance the following: Your account leaves out any reference to the abstract universal *ellipticity*; you describe what goes on as if ellipticity was irrelevant, as if it made no difference to what happens. But the truth of the matter is the ellipticity is the core of the whole process of sorting by this

machine; it is precisely the fact and amount of ellipticity of a particular tile that makes the difference between the two grossly different outcomes of a test—to be 'accepted' or to be 'rejected' by the machine. Evaluating ellipticity is what the machine does, and you have left that out entirely. How can your account be complete, when it treats the abstract universal *ellipticity* as an irrelevancy, as something one can mention or not as he pleases, since it makes no difference to what happens?

In order to decide how much truth there is in this objection, we must first explicate what it means to say that the fact of ellipticity is treated by the micro-analyst as "making no difference to what happens." The most straightforward explication of the phrase 'such-and-such a factor *makes no difference*' would seem to be that the outcome, given the factor's presence, is indistinguishable in all respects from the outcome that would have eventuated assuming the factor to have been absent. So we unpack the criticism "You say that ellipticity makes no difference" as being an assertion by the critic to the following effect: "The non-mathematical mechanician, by putting forth (as allegedly complete) a causal analysis which makes no mention of ellipticity, thereby implies that if the shape of the testing matrix had been other than elliptical, the results would have been the same as they in fact were. Or, in terms of an individual tile (axes 7.5:8) which was accepted, if that tile had been circular, or had axes in the ratio 5:8, it would nevertheless have been accepted." But of course this is false, and is not being asserted by the mechanician. Not only does the mechanician avoid asserting this false counterfactual; he does not assert anything which implies it. On the contrary, what he says *does* imply a counterfactual contradictory to the one imputed, namely, "If the shape of the sorting machine slot had been other than elliptical, the outcome would have been different, i.e., the tiles which dropped through and the tiles which were cast aside ('rejected') would have been different from the tiles that were in fact passed and cast aside." It is one thing to point out, quite rightly, that the mechanical account of the machine's operations fails to mention something which is true, and which follows necessarily from what was mentioned. In this sense the critic is correct in saying "Something has been left out of the account." But this something which has been left out is not a something which involves any new ontological commitments about the furniture of the world, nor is it something which gets us into trouble with the thesis of mechanical determinism. What *would* involve some sort of additional ontological commitments, and would presumably mean that the causal account of the micro-mechanical determinist was defective (i.e., literally left something out, failed to mention a causally significant property), would be an asserted or implied counterfactual, "The elliptical form as a universal instantialized by this particular machine is irrelevant to what the machine does." But that counterfactual is neither asserted nor implied; on the contrary, its contradictory is implied by the mechanical account, even though that account does not use the word "ellipse" or any short synonym thereof. What the account does contain is the large conjunction of sentences giving coordinates of points on the edge, and these coordinates satisfy the equation of an ellipse. Putting aside the necessary refinements of tolerance, physical discontinuity, etc., the mechanician's clumsy conjunction of statements entails (given

the definition of an ellipse) that "ellipticity makes a difference in what happens." What more does the critic want?

It is difficult to summarize the argument presented in the preceding pages. There are five related lines of thought. First, I warn against the temptation to identify "determining factors" in human belief with "non-rational" (e.g., Freudian, Marxian) factors, emphasizing that we have all learned logic as well as other things and that "to think logically" is part of our cerebral computer's programming. Second, I hold that *logical* categories are unavoidable for the psychologist who wishes to deal with human behavior. Third, I argue that the Platonic universals of logic (e.g., Rule of Detachment, *modus tollens*, *dictum de omni et nullo*) are physically modeled by certain subsystems of the human brain, so that—absent countervailing *non*rational forces of Freudian or Marxian type—it tends (statistically) to "think rationally." Fourth, I hold that when a physical system models a calculus, a knower K_c who understand the calculus and knows how to derive theorems within it will have a genuine cognitive edge over a less well-informed knower K_m who knows *everything* K_c knows about the machine's parts + arrangement + laws of mechanics, but who lacks K_c's expertise with the calculus. This genuine cognitive edge is literally *physical* in its content, inasmuch as K_c can actually make correct predictions about the future movements of the machine which K_m cannot make. Fifth, I argue that even if a complete physicalistic micro-causal account might be given of human, rational decision-making—an account that contains no explicit reference to *reasons*—the truth of such a complete micro-causal account would be compatible with a "molar" account truly asserting that *reasons decisively influenced the choice.*

As I stated at the beginning, none of these arguments is intended to show that complete psychological determinism obtains, a thesis which I consider open on present evidence, and unnecessary for the current conducting of psychological research.

References

Ayllon T., & Azrin, N. (1968). *The token economy: A motivational system for therapy and rehabilitation*. New York: Appleton-Century-Crofts.

Bennett, J. (1964). *Rationality*. London: Routledge & Kegan Paul.

Bohr, N. (1934). *Atomic theory and the description of nature*. New York: Macmillan.

Brodbeck, M. (1963) Meaning and action. *Philosophy of Science*, *30*, 309-324.

Brodbeck, M. (1966). Mental and physical: Identity versus sameness. In P. K. Feyerabend & G. Maxwell (eds.), *Mind, matter, and method* (pp. 40-58). Minneapolis, MN: University of Minnesota Press.

Carnap, R. (1934). *Logical syntax of language*. (Trans. Amethe Smeaton). New York: Humanities.

Carnap, R. (1939). *Foundations of logic and mathematics*. vol. 1, no. 3 of *International Encyclopedia of Unified Science* (ed., O. Neurath). Chicago, IL: University of Chicago Press.

Catania, A. C. (ed.) (1968). *Contemporary research in operant behavior*. Glenview, IL: Scott, Foresman.

Cattell, R. B. (1946). *Description and measurement of personality*. Yonkers-on-Hudson, NY: World.

Cattell, R. B. (1950). *Personality*. New York: McGraw-Hill.

Chisholm, R. & Sellars, W. (1958). Intentionality and the mental. In H. Feigl, M. Scriven, & G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science, vol. II* (Appendix, pp. 507-539). Minneapolis, MN: University of Minnesota Press.

Chomsky, N. (1959), Review of *Verbal behavior*, by B. F. Skinner. *Language*, *35*, 26-58.

Eccles, J. C. (1951) Hypotheses relating to the brain-mind problem," *Nature*, *168*, 53-57.

Eccles, J. C. (1953). *The neurophysiological basis of mind*. Oxford: Oxford University Press.

Eddington, A. S. (1929). *The nature of the physical world*. New York: Macmillan.

Eddington, A. S. (1935). *New pathways in science*. New York: Macmillan.

Eddington, A. S. (1939). *The Philosophy of Physical Science*. (Cambridge: Cambridge University Press.

Feigl, H. (1967). *The "Mental" and the "Physical*": *The Essay and* a *Postscript* Minneapolis, MN: University of Minnesota Press.

Feigl, H. & Meehl, P. E. [1974]. The determinism-freedom and body-mind problems. In P. A. Schilpp (ed.), *The philosophy of Karl Popper* (pp. 520-559) LaSalle, IL: Open Court.

Ferster C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.

Freud, S. (1957). The unconscious. In Strachey (ed.), *Standard edition of the complete psychological works of Sigmund Freud*, vol. 14 (pp. 159-216). London: Macmillan. (Original publication 1915)

Grünbaum, A. (1953). Causality and the science of human behavior. In H. Feigl & M. Brodbeck (eds.), *Readings in the Philosophy of Science*. New York: Appleton-Century-Crofts.

Hempel, C. G. (1959/1965). The logic of functional analysis. In L. Gross (ed.), *Symposium on sociological theory*. New York: Harper. Reprinted with alterations in C. G. Hempel, *Aspects of scientific explanation* (pp. 297-330). New York: Free Press, 1965.

Hempel, C. G. (1965). The concept of rationality and the logic of explanation by reasons. In Hempel, *Aspects of scientific explanation* (pp. 463-486). New York: Free Press.

Holland, J. G., & Skinner, B. F. (1961). *The analysis of behavior*. (New York: McGraw-Hill.

Honig, W. K. (ed.) (1966). *Operant behavior: Areas of research and application.* New York: Appleton-Century-Crofts.

Hull, C. L. (1943) *Principles of behavior*. New York: Appleton-Century-Crofts.

Hunt, E. (1968). Computer simulation: Artificial intelligence studies and their relevance to psychology. In P. R. Farnsworth, M. R. Rosenzweig, & J. T. Polefka (eds.), *Annual Review of Psychology*, *19*, 135-168.

Jordan, P. (1955). *Science and the course of history*. New Haven, Conn.: Yale University Press.

Krasner L., & Ullmann, L. P. (eds.) (1965). *Research in behavior modification*. New York: Holt, Rinehart & Winston.

Littman, R. A., & Rosen, E. (1950). Molar and molecular. *Psychological Review, 57*, 58-65.

London, I. D. (1952). Quantum biology and psychology. *Journal of General Psychology. 46*, 123-149.

MacCorquodale, K. (1970). On Chomsky's "Review of Skinner's *Verbal behavior*." *Journal of the Experimental Analysis of Behavior*, *13*, 83-99.

MacCorquodale, K., & Meehl, P. E. (1954) Edward C. Tolman. In W. K. Estes *et al*., *Modern learning theory*. New York: Appleton-Century-Crofts.

Macklin, R. (1968). Doing and happening. *Review of Metaphysics*, *22*, 246-261.

Macklin, R. (1969). Action, causality, and teleology. *British Journal for the Philosophy of Science*, *19*, 301-316.

Meehl, P. E. (1966) The compleat autocerebroscopist: A thought-experiment on Professor Feigl's mind-body identity thesis. In P. K. Feyerabend & G. Maxwell (eds.), *Mind, matter, and method* (pp. 103-180). Minneapolis, MN: University of Minnesota Press.

Meehl, P. E., Klann, R., Schmieding, A, Breimeier, K, & Schroeder-Slomann, S. (1958). *What, then, is man?* St. Louis, MO: Concordia.

Murray, H. A. (1938). *Explorations in personality*. New York: Wiley.

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

Newell, A., Shaw, J. C., & Simon, H. A. (1957). Empirical explorations with the logic theory machine. *Proceedings of the Western Joint Computer Conference*, *11*, 218-230.

Newell, A., Shaw, J. C., & Simon, H. A. (January 13, 1958). Note: Improvement in the proof of a theorem in the elementary propositional calculus. C.I.P. Working Paper no. 8 (in ditto form, available from Dr. Simon).

Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a general problem solving program. *Proceedings of the International Conference on Information Processing* (pp. 256-264). Paris: UNESCO.

Pap, A. (1962). *An introduction to the philosophy of science*. New York: Free Press,

Pirenne M. H., & Marriott, F. H. C. (1959). The quantum theory of light and the psychophysiology of vision. In S. Koch (ed.), *Psychology, a study of a science,* vol. I: *Sensory, Perceptual and Physiological Formulations* (pp. 288-361). New York: McGraw-Hill.

Platt, J. R. (1966). *The step to man*. New York: Wiley.

Popper, K. R. (1962). "Why are the calculi of logic and arithmetic applicable to reality?" In Popper, *Conjectures and refutations*. New York and London: Basic Books.

Popper, K. R. (1966). "Of Clouds and Clocks: An Approach to the Problem of Rationality and the Freedom of Man," the Arthur Holly Compton Memorial Lecture presented at Washington University, April 21, 1965. St. Louis, MO: Washington University, 1966.

Popper, K. R. (1968). Epistemology without a knowing subject. In B. Van Rootselaar & J. F. Staal (eds.), *Logic, methodology and philosophy of science*, vol. III (pp. 333-373). Amsterdam: North-Holland.

Ratliff, F. (1962). Some interrelations among physics, physiology, and psychology in the study of vision. In S. Koch (ed.). *Psychology, a study of a science,* vol. 4: *Biologically oriented fields*. New York: McGraw-Hill.

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

Samuels, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*, 210-229.

Sayre, K. (1965). *Recognition: A study in artificial intelligence*. Notre Dame, IN: University of Notre Dame Press.

Schroedinger, E. (1951). *Science and humanism* (Cambridge: Cambridge University Press.

Sellars, W. (1947a). Pure pragmatics and epistemology. *Philosophy of Science*, *14*, 181-202.

Sellars, W. (1947b). Epistemology and the new way of words," *Journal of Philosophy. 44*, 645-660.

Sellars, W. (1948). Realism and the new way of words. *Philosophy and Phenomenological Research*, *8*, 601-634.

Sellars, W. (1952). Mind, meaning, and behavior. *Philosophical Studies*, *3*, 83-94.

Sellars, W. (1954) Some reflections on language games. *Philosophy of Science, 21*, 204-228

Sellars, W. (1953). A semantical solution of the mind-body problem. *Methodos, 5*, 45-85.

Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science, vol. I* (pp. 253-329). Minneapolis, MN: University of Minnesota Press. Also in *Science, perception, and reality*. New York: Humanities, 1963.

Sellars, W. (1963). Empiricism and abstract entities. In P. A. Schilpp (ed.), *The philosophy of Rudolf Carnap* (pp. 431-468). LaSalle, IL: Open Court.

Sellars, W. (1966). Thought and action. In K. Lehrer (ed.), *Freedom and determinism*. New York: Random House.

Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1957). *Verbal behavior* (New York: Appleton-Century-Crofts.

Skinner, B. F. (1951). *Science and human behavior.* Cambridge, MA: Harvard University Press.

Skinner, B. F. (1961). *Cumulative record*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1968). *The technology of teaching*. New York Appleton-Century-Crofts.

Stebbing, L. S. (1937). Causality and human freedom. In L. Stebbing, *Philosophy and the physicists*. London: Methuen.

Tolman, E. C. (1932). *Purposive behavior in animals and men*. New York: Appleton-Century-Crofts.

Tolman, E. C. (1951). *Collected papers in psychology*. Berkeley, CA: University of California Press.

Ullmann L. P., & Krasner, L. (eds.) (1965). *Case studies in behavior modification*. New York: Holt, Rinehart & Winston.

Ulrich, R., Stachnik, T., & Mabry, J. (eds.) (1966). *Control of human behavior*. Glenview, IL: Scott, Foresman.

Verhave, T. (ed.) (1966). *The experimental analysis of behavior: Selected readings*. New York: Appleton-Century-Crofts.

von Wright, G. H. (1968). The logic of practical discourse. In R. Klibansky (ed.), *Contemporary philosophy* (pp. 141-167). Florence: Nuova Italia.

Weber, M. (1949) A critique of Eduard Meyer's methodological views (1905). In M. Weber, *Methodology of the social sciences*, (ed. E. A. Shils & H. A. Finch). New York: Free Press.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: K. Paul, Trench, Trubner.

Woodworth, R. S. (1938). *Experimental psychology*. New York: Holt.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning," *Journal of Experimental Psychology*, *18*, 451-460.

pdf by LJY, May 2016