Feigl, H. & Meehl, P. E. (1974). The determinism-freedom and mind-body problems. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 520-559). LaSalle, IL: Open Court.
Reprinted in P. E. Meehl, *Selected philosophical and methodological papers* (pp. 97-135; C. A. Anderson and K. Gunderson, Eds.). Minneapolis: University of Minnesota Press, 1991.

Chapter 14

Herbert Feigl and *Paul E. Meehl*

THE DETERMINISM-FREEDOM AND BODY-MIND PROBLEMS

Our cherished friend, Sir Karl, has very penetratingly and challengingly dealt in several essays[1] with two of the most difficult and controversial issues of modern philosophy and science. In accordance with Popper's own designations we shall speak of "Compton's problem" and "Descartes's problem". Compton's problem is how to account for free choice and genuine (artistic or scientific) creativity; Descartes's problem concerns the relation of the mental to the physical. These problems are closely related, and both are viewed by Popper in the light of modern physics, biology, psychology and of the theory of language. Although we do not pretend to know of any definitive solutions of either problem, and although we tend to agree with several important points made by Popper, we propose to submit to him a number of critical reflections. What we hope to show is that the "plastic" (or "cloud-like") control stressed by Popper does not necessarily require a basic indeterminism; and that the role of meanings and reasons in the representative and argumentative functions of language does not necessarily imply a dualism of mind and body.

I

Popper's "nightmare of determinism"—very much like the dread of the "block universe" of William James—seems to us to rest on identifying determinism with strict predictability. We can hardly believe that Popper regards these terms as synonymous. If so, this could only be due to an unrelinquished remnant of positivistic thinking in the keenest and most outstanding contemporary critic of positivism. It seems quite unlikely that Popper, for once, and quite contrary to his general enlightened, has fallen victim to a verificationist prejudice. This would be inconsistent with his pronouncements in other places on the *metaphysical,* i.e., untestable nature

of the doctrine of determinism. In any case, we think that complete predetermination becomes a "nightmare" only if one assumes that some sort of Laplacean World Formula could ever be produced by scientists. Such a World Formula would then—inserting the total set of specific numerical values for the initial and boundary conditions—furnish a World Calendar. In such a calendar one could look up one's own future, the future of mankind and of our planet in general, etc., in perfect detail; and if the basic laws are temporally symmetrical (as they are at least in classical physics), one could also reconstruct any phase of the historical, prehistorical, or cosmological past. It is the attainability of a World Formula thus understood which leads to absurdly paradoxical and abhorrently unpalatable consequences. Fatalism would then seem the only possible attitude. Any avoidance reaction to predictions of unpleasant or disastrous events would itself have to be derivable from the World Formula, and the well-known phenomenon of a self-annulling prediction would be logically impossible—as long as we assume the World Formula to be correct. Precise and detailed predictions of future works of art, of scientific theories, of the rise or decline of civilizations, etc., would all be attainable in terms of an accurate physicalistic description of every and all spatiotemporal events-indeed including the "putting of ink marks on paper"—be it a future composer's symphonic score, or the manuscript of a great mathematician or theoretical physicist of the year 2500! If such precise predictions were possible, then we might have full knowledge of scientific theories, or achievements of works of art, long before they were—respectively—invented or created. This is surely as logically incoherent an idea as is that of H. G. Wells's time machine! Furthermore, if the World Formula permitted the prediction of a fatal automobile accident which someone is to suffer at a definite place and a precise date—in the future, then he could not succeed in conveniently staying at home, and thus avoid the disaster. By a curious and intricate concatenation of circumstances he would nevertheless be "destined" to die in precisely the predicted circumstances, and in the specified space-time region. If the World Calendar contained a prediction of even some highly desired event, one could not—merely for the sake of disproving the World Formula—change the course of events such that the predicted one would not take place.

Considerations such as these have often been adduced as a reductio ad absurdum of determinism. But it should be clear that it is rather the attainability of the World Formula which the foregoing arguments refute. Now, as is agreed on all hands, the idea of the World Formula is to be understood as a logical conjunction of *three* propositions: (1) The doctrine of the deterministic form of all basic natural laws. (2) The precise, complete (and simultaneous) ascertainability of all initial and boundary conditions. (3) The mathematical feasibility of the hopelessly complex computations

necessary for precise and complete predictions (or retrodictions). Now, as no one knows better than Popper (though this is really a matter of the most elementary propositional logic), if a conjunction of several independent propositions entails a false or absurd conclusion, not every one of the conjuncts is (necessarily) false. In the empirical sciences there are experimental or statistical methods to pinpoint the "culprit" (or culprits) among the conjoined premises. In our problem this is obviously impossible. Nevertheless, there are excellent reasons for regarding propositions (2) and (3) as false at any rate; thus leaving the hypothesis of determinism at least open for further consideration.

We should make it quite clear immediately that neither in the above remarks, nor in what follows, are we pleading for the doctrines of determinism. We consider it quite possible that the indeterminism of present quantum mechanics (or something akin to it) may never be overcome. On the other hand, it is conceivable that a future theory regarding a further substratum of micro-microevents might have deterministic form and be nonemergentist. The question as to what form—deterministic or statistical—the "rock bottom" laws of the universe have, is indeed never conclusively decidable; and this for the simple reason that there is not and could not be a criterion by which to recognize the "rock bottom" of nature. As David Bohm has suggested, it is logically conceivable that our universe may be "infinitely deep"—layer upon layer without end. Hence, the only sensible question to ask is whether at a given stage of scientific (experimental and theoretical) investigation the (on that level) basic laws are strictly causal (deterministic) or statistical (probabilistic). Hence, while we agree with Popper that neither the doctrine of determinism nor that of indeterminism is conclusively decidable,[2] we think that empirical evidence, that is reasons, can be adduced that justify a tentative endorsement of one or the other doctrine. Surely, if the triumphant successes of classical mechanics, and of the nineteenth-century field theories had continued to furnish adequate explanations in every domain of the empirical sciences, determinism would have remained a highly plausible frame hypothesis.

Since it is imperative for our conception of determinism to separate it quite radically from all references to prediction and predictability, perhaps an explicit formulation of "strict lawfulness" is needed. We suggest the following definition: Any event in the world (as described in the four-dimensional representation) instantiates a nomological proposition—i.e., either a basic or a derived law. Thus understood, the frame hypothesis of determinism is not hopelessly untestable. Indeed, it was the unexpected development of quantum physics in our century that for the first time cast *serious* doubt upon the determinism doctrine. But before we turn to the implications of indeterminism for "Compton's problem", let us show that the two other conjuncts in the World

Formula doctrine are much more vulnerable than the idea of determinism. As Popper himself has shown,[3] even under the presupposition of the determinism of classical physics, there is a fundamental (we should think set-theoretical) impossibility in ascertaining and recording the initial and boundary conditions of a closed system. As long as the observing-measuring instrument and/or observer is part of the system under scrutiny, not all of the system can be "mapped" on to part of itself. This would indeed seem a *logical* impossibility as regards proposition (2), i.e., of the full ascertainability of the total set of conditions. To this might be added the basic physical impossibility of knowing about incoming "inputs" before they have arrived. According to the extremely well-corroborated principles of the special theory of relativity, any "messages", "information" (really any sort of propagation of causal influences) cannot occur at a velocity greater than the speed of electromagnetic waves. Hence—strictly speaking—there can be only ex post facto explanations of events—but no rigorous and completely reliable predictions of them. Only the events in Minkowski's "passive light cone of the past" can be adduced for predictions. Any events "outside that cone" can become known only after a certain time has elapsed. While Bergson (who never quite understood Einstein's theory) did not base his early pronouncement of free will (in the sense of prior unpredictability of action) on relativistic principles, he was, nevertheless, correct in claiming that most actions could be causally explained only after they had occurred. Hence, we agree with Popper that even on the basis of the theories of classical physics (including the special and general theories of relativity) complete and precise prediction is logically and physically impossible.

We disagree with Popper, however, if he views these impossibilities of prediction as arguments against determinism. (They are, we insist, decisive arguments only against the feasibility of the World Formula.) We view the doctrine of determinism as meaningful, coherent, though, of course, only very inconclusively testable. Perhaps this merely reveals that—Popper's critique to the contrary notwithstanding—we are not as radically anti-inductivistic as he is. The very fact that one could *reason* from the successes of classical physics to the (then) plausible assumption of determinism; and that the laws of classical physics are deterministic in their mathematical formulation—all this indicates that one can understand the meaning of determinism even if this doctrine is operationally and practically inconsequential. We do understand the counterfactual proposition: "If the totality of initial and boundary conditions could be precisely ascertained; and if the laws are deterministic; and if the immensely complex computations could be accomplished, then every event in the universe could be exactly predicted (or retrodicted)". This is the conjunction of our three propositions all over again. The idea of determinism by itself need not be wrong. It could—as it has for a long time—still serve as

the guiding maxim of scientific research. It is merely "sour grapes" policy of scientists or philosophers of science, if they maintain that "statistical causality" is just as good as fully deterministic lawfulness. Einstein's well-known opposition to regarding quantum mechanics as complete, emphatically expressed his conviction that nature "at rockbottom" (?) is strictly deterministic. Having succeeded twice (viz., in his special and general theory of relativity) in a "geometrization" of physics, he hoped to succeed once more—in his attempts towards a unified field theory—to design a new fundamental account, and thus to eliminate the idea of "absolute chance" conclusively from all future science. To his great disappointment he failed in this ambitious endeavor.

On a much lower level of sophistication this "faith" in determinism is clearly exemplified by the traditional view of the games of chance. Which way spinning coins, or dice, or roulette balls, will come to rest is considered a matter of *relative chance.* The thought here is that the—no matter how complex and delicate—initial and boundary conditions strictly determine the outcome. And if we only knew their precise total constellation, we could predict the outcomes with full certainty and precision. The very concept of relative chance then presupposes determinism. Chance in this sense is relative in a twofold way: (1) Relative to our knowledge (or ignorance) the event in question is not precisely predictable. (2) The many factors or conditions relevant for the outcome have a high degree of causal independence from each other. This is the way we customarily view "accidents" or "coincidences", be it a cosmic collision of stars, or an automobile accident. Essentially this rests on the causally contingent relations of events that occur at a distance from one another and are (roughly) simultaneous. In the well known four-dimensional representations this is indicated by the crossing of world-lines. In the games of chance (and as we shall see in a moment, also in the kinetics of molecules), there is the further very important aspect of the most "delicate" dependence of outcomes upon extremely small variations in the initial conditions. As to whether, for example, a coin will come up heads or tails; or as to whether a ball on the Galton board will turn left or right after impinging on a given nail, will often depend—according to classical mechanics—on minimal differences in original positions, speeds, friction encountered, etc. Surely, here is one more extremely powerful argument for the impossibility of precise and certain predictions even under the presupposition of strict determinism.

Ascertaining by measurement of the initial conditions would have to be *completely* exact, for even infinitesimally small differences in the causal antecedents may result (by various amplification processes) in enormous differences in the ensuing effects. In other words, continuous variations in the initial conditions can bring about discontinuous changes in highly relevant features of the causal results. The ball coming down from a nail on the Galton

board may finally reach one electric relay that brings about the ringing of a bell; or it may trigger off another relay, and thereby cause the explosion of a bomb. And if the initial conditions were such that the ball were coming down exactly vertically upon the nail, it should—after a few dampened bouncings—come to rest, in unstable equilibrium right on top of the nail. This latter—unheard of—event would, however, be excluded by the random (heat) motion of the molecules, both in the surface of the ball and in that of the nail. Since all measurements—even from the classical-deterministic point of view—are achieved by using macroinstruments, and since molecular motion (at least at temperatures above absolute zero) is unavoidable, there is even in the classical theory an ineluctable inaccuracy (i.e., an element of chance) in the results of even the most refined measurements.

There is no need, we are sure, for any elaborate argument regarding the overwhelming complexity of the calculations that would be required for precise predictions for any but the most simple (and idealized) systems and their "development". As all the world knows, this is the *raison d'être* of the statistical approach—be it in the games of chance, in matters of insurance, or in the kinetic theories of molecular motions. The sort of system of equations that would be required for detailed and precise predictions would obviously involve difficulties far surpassing those of the astronomical *n*-body problem. It is, moreover, quite conceivable that the pure mathematics of such calculations might run up against insoluble mathematical problems—in the sense of the undecidability discovered by Kurt Gödel.

All these considerations, we repeat, do not however establish conclusive argument against determinism as such. To illustrate in terms of Popper's own example, the "behavior" of soap bubbles can be explained, as Ludwig Boltzmann might well have claimed, within a deterministic point of view. To be sure, precise predictions (e.g., of the moment of the bubble's bursting) are impossible—for the reasons already mentioned. Yet, according to Boltzmann's convictions, the kinetics of the air molecules inside and outside the bubble, as well as of the molecules constituting the thin "skin" of the bubble, might be construed in the light of Newtonian determinism. Boltzmann might have said that we have very good reasons for assuming such an "underlying" determinism, in spite of the hopelessness of detailed and exact predictability. Classical determinism (involving, of course, only concepts of *relative* chance) thus would provide a perfectly intelligible basis for the "plastic control" that Popper emphasizes. As he admits, "clouds" may well be "clocks" in this sense, after all. Very much, however, depends on what one means here by "in this sense". As Popper rightly sees, there is a continuum of degrees that lies between extreme cases of "clouds" and extreme cases of "clocks".

Analogous considerations apply to the differences of organisms and machines. Admittedly, there are tremendous differences between a simple

mechanical machine (such as a system of levers and pulleys, or of cog wheels) and even the simplest organism, such as a protozoan. Perhaps even a difference in kind must be admitted between noncybernetic mechanisms and mechanism involving self-regulation. We must not prejudge the issue by any simplistic "man a machine" conceptions. We are inclined to think that in any case machines made of hardware (vacuum tubes, transistors, wires, transmission gears) no matter how ingeniously constructed, will, at best, simulate (or perhaps even "outdo") only certain segments of human activities. Very likely only a structure composed of the proteins of the nervous system can function in the "creative" manner of the highest human achievements. An anecdote about a conversation between a brilliant avant-garde engineer and a great musician poignantly illustrates the issue. The engineer said: "Now we can build machines that compose works of music". The musician replied: "Then you should also build some machines that will appreciate that sort of music". The point is well taken. Electronic "music" of the "aleatory" type is one of the atrocities of our age of technology. We agree wholeheartedly with Popper that human creativity cannot be explained on the basis of a "clock" (simple machine) theory of cerebral functioning.[4] We also agree that the combination of a machine with a built-in randomizer will not suffice either. But it has become increasingly clear that systems with "organismic" structures, i.e., involving Gestalt and cybernetic features, may well exemplify the typically "teleological" aspects of a great variety of physiological processes. This, together with whatever "randomizing" factors may be at play, could go a long way toward a deterministic explanation of the delicately attuned, but never fully predictable biological and psychological functions and occurrences.

As we understand Popper's views (especially in *Of Clouds and Clocks),* he rejects determinism in order to make room for an (unfortunately none too explicitly formulated) *emergentism.* Here again it seems imperative to us to distinguish between emergence in the sense of unpredictability, and emergence as involving a basic indeterminism. As has often been pointed out,[5] the impossibility of prediction in the cases of emergent novelty does not necessarily imply a denial of determinism. The familiar examples of the impossibility of predicting the properties of complexes (e.g., chemical compounds, organisms, social groups) on the basis of the properties of their constituent parts or components (e.g., chemical elements, cells, individual persons), do not establish an argument against the possibility of deterministic theories that would explain the properties of the "organic wholes". The arguments for "genuine emergence", along with the arguments for ("irreducible") holism appear plausible only as long as they are couched in *epistemic* terms. That is to say, if the properties of the parts or components have been explored only in "splendid isolation", there would hardly be any

clue as to what sort of wholes (compounds, organisms, etc.) they might form when brought into interactions with each other. This is strikingly true even on the most elementary level of modern microphysics. A study of the behavior of "free" electrons (as in the cathode rays) would never suggest their "configurational behavior" in the context of even the simplest atoms. Pauli's exclusion principle (which is an independent postulate of quantum mechanics), may be considered a fundamental composition law; just as in classical mechanics the law of vectorial addition (i.e., the well-known parallelogram of forces) is a basic empirical law, logically independent of the other Newtonian laws of motion and of gravitation. Just how far we have to explore more and more complex situations, constellations and configurations in order to "glean" (pardon the "inductivism"!) the total set of laws sufficient for purely computational (i.e., mathematical, including geometrical) inference of the regularities of complexes (wholes) cannot be decided in an a priori manner. By and large, the extant evidence encourages the view that chemistry is in principle reducible to atomic and quantum physics; and that biology (especially since the recent advances in the knowledge of the molecular structure of the nucleic acids) is equally—in principle(!)—reducible to modern physics. (This is not to deny that there are still grave unsolved problems in connection with the specifics of evolution, and other biological problems.)

The doctrine of "genuine emergence" is incompatible with the sort of reductionism just sketched only if new sorts of regularity were to crop up indefinitely at levels of higher complexity, and if these new regularities were absolutely underivable from the laws of some level of lower complexity. As we see it, there is no reason for such pessimism about the possibility of unitary scientific explanation. Quite to the contrary, there are important instances of the prediction of novel features on the basis of theories that were formulated *before* the resultant features of "compounded wholes" had been empirically known: Consider Mendeleev's predictions of new chemical elements, their properties, and the properties of their compounds; or the prediction of (artificial) fission and fusion of atomic nuclei; predictions of the phenotypical characteristics of organisms on the basis of the theory of genes, etc.

Even if such striking feats of prediction had not been achieved, the very possibility of *theories* from which the properties of "wholes" can be derived, can certainly not be denied on a priori grounds; and, as just indicated, that possibility is becoming increasingly plausible in the light of empirical evidence. This is at least one type of argument for the compatibility of (epistemic) emergence with (theoretical or "ontological") determinism. And even disregarding high-level theories, the conditions and consequences of "emergence" could be stated in the form of empirical laws. Once the resultants of compounding processes have been observed, laws formulating

antecedents as causally implying their consequents would be subject to corroboration. Of course, these laws may have deterministic form (as in most cases of chemical compounding) or they may be statistical (as in genetics). If they are ineluctably statistical, then a certain measure of indeterminacy would be characteristic of emergence. If they are deterministic, then one could always (retrospectively) formulate the regularities in terms of dispositional properties of the components. (For example, sodium and chlorine are "apt" to combine into ordinary salt.)

In general, the logical situation in regard to emergence seems to be as follows: Only if the concepts of the theories (or laws) which are to be reduced (i.e., derived, explained) are explicitly definable in terms of the concepts of the reducing theory, can the reduction (derivation) be accomplished. But, since this is merely a necessary and not a sufficient condition, emergence in the sense of nonderivability may yet obtain, even if the definability condition is fulfilled. Hence, we regard the problem of emergence as an empirical question.

It is generally agreed that, e.g., laws of electromagnetism are not derivable from the laws of classical mechanics. Thus one could say that the phenomena of electromagnetism are "emergent" with respect to mechanical phenomena. It is even tempting to say that a sort of "interaction" between the electromagnetic forces and the (mechanical) masses takes place—as in the cases of "ponderomotoric" electric forces or in the phenomena of light pressure. But it is highly questionable as to whether an analogous interaction may be assumed between mental states (once they have emerged—be it in the course of phylogenetic evolution—or in the course of ontogenetic development) and the respective nervous systems. The resistance to this sort of doctrine among natural scientists and physicalistically oriented philosophers is not merely due to its "spookiness". The idea that "immaterial" mental factors somehow interact with, or intervene in, the physiological processes is not necessarily untestable, let alone meaningless. Of course, one would want to know more precisely what is meant by "immaterial" or "nonphysical". But once that is at least in outline specified, the remaining question is whether a hypothesis of this kind is needed. If some sort of nonphysical agent (let us call it "psychoid") is required in order to account for free choice and creativity, how are we to square this with all that is known about the factors that are relevant for the formation of the character, personality and intelligence of human beings? Hereditary constitution as based on the genetic makeup, together with all the influences of maturation, education (positive and/or negative reinforcements in the physical and social environments), emulation of "father figures" (or oedipal reactions against them), etc., seems to offer a sufficient basis—and one that may well be—in principle—susceptible of a physicalistic account.

The repudiation of dualistic-interactionistic explanations can be understood, if not justified, in the light of all the accumulated scientific evidence that speaks for a view of causality that is "closed" in the physical world, i.e., not open to "miraculous" interventions by nonphysical causes. The retreat of animism and vitalism is surely due to the ever growing scope of physical explanations. Driesch, the well-known vitalist, some forty years ago postulated entelechies which, though not in space, were assumed to "act into space". And anyone who assumes that the indeterminism of modern physics allows for "loopholes" or a "leeway" for the intervention of nonphysical forces, proposes in our century a doctrine homologous to the one proposed by Descartes in the seventeenth century. Descartes had to postulate a "breach" in the mechanical laws in the case of mental-physical interaction. Similarly, if there is to be anything testable in the twentieth-century idea that quantum indeterminism allows for (at least slight) influences of immaterial mental states upon the ongoing atomic processes in human brains, some sort of "breach" in, or deviation from, the statistical regularities would have to be assumed. Perhaps this sort of "violation" could be made more palatable by assuming that the local frequencies of quantum transitions (or the like) would be in keeping with the physical theory, but that a sort of "patterning" in the firing of neurons takes place which is not derivable from physical theory alone, and would thus require the intervention of a "psychoid". Hence, both Descartes and Compton alike make assumptions, which—if testable—deviate from what the physics of their respective times (seventeenth and twentieth centuries) would imply.

Naturally, all the remarks just made are merely the results of logical analyses. It does not seem likely that an experimental test will be forthcoming in the foreseeable future. Physicalistically oriented thinkers will insist that Compton merely repeated Descartes's mistaken reasoning on the more sophisticated level of twentieth-century science. Thinkers opposed to physicalism find this type of interactionism not so terribly "strange" after all. What we intended to point out is merely that some "tampering" with physical laws is unavoidable—even if the laws are partly (or largely) statistical, as they are in quantum mechanics.

Before turning to the role of meaning, rules and norms in human behavior—and thus to "Descartes's problem", a few more remarks on free will are in order. Let us candidly admit that we, too, can feel the sting of this perennial puzzle. Surely, if we human beings are solely the products of "nature and nurture", i.e., of our inherited constitutions and all the influences that have impinged upon us ever since the fertilization of the ovum from which we developed—how are we to understand the sentence asserting "in the given life situation we *could have acted differently* from the way we actually did". G. E. Moore's well-known answer "we could have acted differently, had

we wanted to" merely shifts the vexing question to the freedom of "wanting". Some misguided thinkers even went so far as to question the very idea of moral responsibility by pointing out that we did not make or choose our character or personality. The current revolt against the Hume-Mill-Schlick-Hobart "dissolution" of the free will versus determinism issue indicates that perhaps we latter-day "compatibilists" should add a few further clarifying observations to the old issue. Surely it must be admitted that, for example, Schlick, despite the great clarity of his thought on the issue, did, in part, mis-diagnose its roots. Schlick was very likely wrong in thinking that the trouble stems from confusing descriptive (natural) law with prescriptive (judicial) law. We are not aware of any important philosopher who committed this error. But the "compatibilist" school of thought seems to us to have provided some very important and very illuminating considerations. There is first, the obvious distinction between voluntary and compelled action. Both can be understood on the basis of deterministic assumptions. Even purely behaviorally there are (almost in the sense of physics) "degrees of freedom" with respect to the environmental context of our voluntary action. The ball on the inclined plane is "bound to" roll down. But the rabbit on the hillside might well go up or sideways. And a human being on that same hillside may sing a song, write a letter or poem, do gymnastic exercises, or eat his lunch, etc. While this is only a beginning in the explication of "freedom *of* choice", it is an indispensable first step. It could be formulated in terms of the comparative predominance of the internal (intradermal) factors over the external (extradermal) ones in the causal determination of behavior. However, to define "free choice" merely negatively by the absence of compulsion, coercion or constraint, though stating necessary conditions, is not by itself sufficient. Positively, there is the undeniable fact that in voluntary action, we are the doers of our deeds. Our character and personality, and our desires, express themselves in our deliberations, decisions and actions. And although the causal account cannot as yet be given in all of its detail, it should be clear that we are essential links in the causal chains of such voluntary actions. Any choice we make after surveying and contemplating the possible avenues of actions reflects our character, our momentary attitudes, sentiments, and moods. Of course, any precise and fully reliable predictions are precarious, if not impossible. Nevertheless, even the very incomplete and inexact knowledge we have acquired of our relatives, friends, and close acquaintances, often enables us to foretell, with a high success frequency, their reactions to various life situations. We know what will please, amuse, or annoy them; we know their preferences—be it in the choice of foods or drinks, or in voting for political candidates. We know that employees practically always cash their salary checks (sooner or later—usually sooner).

The ascription of responsibility for an action necessarily involves (in ad

dition to any moral-normative considerations) causal imputation. In common language this manifests itself even in locutions that have nothing to do with moral judgments. We say, e.g., "the earthquake is responsible for the devastation", "the landslide is responsible for the change in the river's course, etc. The work of a great scientist is *his,* for the simple reason that *he* produced it—never mind in that context what "produced him"! If the music of Beethoven in his late quartets (Op. 127 to Op. 135) "reflects" something of his personality of that period, does this not mean that there is a causal relation between the composer and his work? To be sure, one must not be so naïve as to expect any simple and predictable relation here. Mozart produced some of his most serene and happy-sounding music in periods of great stress—if not distress. Nevertheless, there is no reason for abandoning the scientific approach which—no matter how qualified by probabilistic considerations—is after all *causal* analysis. And causal analysis, in modern interpretation, amounts to nothing else but the search for relevant variables and their lawful functional relationships. Surely coincidences of all sorts may have to be taken into account. Beethoven is reported to have been stimulated occasionally by hearing a poplar tune whistled by someone in the streets of Vienna. Is this *fundamentally* different from the changes brought about in planetary motions by an "accidental" approach of a comet?

Summing up, we submit that what we know, cherish, and designate as "free choice" is not only compatible with determinism, but actually presupposes a large measure of it for the processes constituting human actions. As Popper rightly stresses, nothing is gained by the assumption of absolute chance—i.e., of completely uncaused action. What, then, could indeterminism do for genuine freedom? All it could do, we recapitulate, is to give us one more (and we admit, of course, fully decisive) argument against the attainability of the World Formula. It seems to us that the "nightmare" of the World Formula is indeed—on all counts—a chimera, a fantasy gone wild. It is in this sense on a par with theological doctrines of predestination. Divine omniscience poses exactly the same problem, and some brilliant theologians (like Jonathan Edwards) have seen quite clearly that this is perfectly compatible with human free will and moral responsibility.

Now let us face the free will perplexity once more in its most poignant form: If strict determinism were true then there are no genuine alternatives for human action. "I could have done otherwise" can only mean that if I had been different (e.g., wiser or better) I would have acted differently. But precisely because human nature is "flexible", i.e., capable of learning from the lessons of experience, it can change from one occasion to the next. This is certainly a most important feature in which human beings (but also many species of animals) differ from *simple* automata whose internal structure is rigidly fixed, and thus not responsive to the "lessons of experience". The sen-

timent of regret, while phenomenologically directed upon the past act ("I wish I had not done it"), finds its pragmatic significance in the resolution to "reform" ("Next time I'll behave differently"). Moral responsibility presupposes responsiveness to the sanctions of society. The severely insane are therefore not held responsible. Does "making an effort" make a difference? It certainly does, at least sometimes. Why not view the human conscience (somewhat as Freud suggested in his "anatomy of the personality") as a subsystem of the person (physicalistically: the organism)? As such the superego is indeed involved in highly subtle interactions with other subsystems (the ego and the id). Hence, if for example, a certain type of education results in the formation of a powerful superego, this may well manifest itself in the type of conduct a person displays.

We repudiate the idea that such free will as we possess is an illusion, arising from the ignorance of the causal conditions of our actions. There are situations in which we know the relevant antecedents quite well, and nevertheless do what we do "of our own free will". For example, we know that we are very fond of our great friend, Sir Karl; hence we quite voluntarily try to write in a manner that will not offend him. Only a detailed, complete foreknowledge such as the chimerical World Formula could provide, would deprive us of our feeling of free choice.

If all of the foregoing considerations do not remove the sting of the free will versus determinism problem, then perhaps only a word of practical wisdom can help. It is morbid to contemplate universal causation while engaged in making decisions in the context of practical urgencies. Fortunately, it is rather difficult, if not psychologically impossible, simultaneously to combine the attitude of the spectator and causal analyst with that of the goal-pursuing agent. (To the complementarity philosophers a word of warning: this has nothing whatever to do—except for a very remote analogy—with the complementarity formulated in the Copenhagen interpretation of quantum mechanics!)

We conclude that the testimony of introspection, as well as the objective observation of behavior at "choice points" quite clearly reveals the efficacy of deliberation and effort. Even if deliberation, preference, choice and action were completely determined by causal antecedents, it is still *free* choice (as contrasted with decisions imposed on us by any form of compulsion—from the normal cases of coercion to the peculiar cases of hypnosis or "brainwashing"). If "classically" oriented scientists—from Laplace to Einstein—after careful consideration of all relevant reasons—came to accept and defend a deterministic point of view, their convictions are "rational" only to the extent that they responded to the available evidence, and with their mastery of a variety of theoretical schemes, in a manner that might well be ultimately explainable in causal terms. This seems basically not different

from the trial and error-elimination process involved in all cases of learning. In fact, it seems to us that (to use Sir Karl's way for formulating the matter) the refutation or the corroboration of theories had better be intelligible in causal-psychological terms, if it is not to consist of "snap judgments".

Although we grant, of course, that in many contexts the words "cause" and "reason" are used (and should be used) in categorically different ways, we submit that there are contexts in which they are practically synonymous. And even where they are not—there are subtle relations between them. This is one of the main points to be discussed in the following section.

II

Turning now to Popper's views regarding Descartes's problem, i.e., the "body-mind" problem, it will be well to point out the primary issues in the cluster of questions that are traditionally and controversially discussed under that heading. We think it useful to distinguish three major parts of that cluster: Sentience, Sapience, and Selfhood. Of these it is *sapience* which is in the foreground, both for Descartes and for Popper. The problems of sapience concern the intellectual capacities and activities of human beings—in relation to the processes occurring in their nervous systems, particularly in the cortices of their brains. Everything that is relevant for perceiving, knowing, reasoning, problem solving, and the like is here comprised under "sapience". "Sentience", by contrast, designates the qualities of immediate experience, and the problem here also is to give a consistent, coherent and scientifically acceptable account of its relation to the neurophysiological process. "Selfhood", in turn designates the "identity" of the human person—among other persons—and as a continuant throughout a span of time. Here, too, there are questions as to the relationships between both the introspective and the common-life descriptions of a person, and the (ultimately physical) scientific account of the organism as the "embodiment of a mind". Popper, most understandably and justifiably, focuses on the problem of sapience. We can best set the stage for our critical examination of his position (that human sapience is incompatible with determinism) by gathering together the chief passages in *Of Clouds and Clocks* which most explicitly assert and argue this incompatibility. Popper writes:[6]

> [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 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. [P. 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 [29]. 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 [30]. [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 does 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]

These passages mention several distinguishable components of human mental life (e.g., our feelings, our desires, our reasonings) some of which are more clearly sapient (= cognitive) than others. Thus, for example, a person's "desire for water" would ordinarily be viewed by the psychologist as a fairly complex state of the organism, including such component aspects as water depletion in the peripheral body tissues, afferent nerve impulses arising from local dryness of the throat and mouth, chemical conditions of the extracellular fluid surrounding and bathing nerve cells in the thirst-specific and drinking control centers of the hypothalamus, a heightened "arousal" of the generalized sort associated with any strong biological drive, selective perceptual sensitization to water-related exteroceptive cues, differential activation of acquired instrumental habits (including verbal ones!) that have been strengthened by water reinforcement in the past, and so on. It seems safe to assume, on present evidence (both scientific and commonsense) that those states of the organism that we ordinarily subsume under such generic state-terms as "desire", "feeling", "motive" and the like are partly sapient, partly sentient, and partly neither. For example, unconscious wishes are, by definition, not sentient, and they are (quasi-) sapient only by a complicated—and still disputed—extension of familiar meanings.[7]

It is not our intention to consider separately each of the distinguishable kinds of mental events to which Sir Karl alludes in these passages. We do not believe this is necessary, even if our limited space permitted it. We suggest that the core philosophical objection is best represented by the "pure case" of human sapience, to wit, *rational inference* by a calm, nonhungry, nonthirsty, sexually satisfied, unfrightened, nonangry scholar, whose regnant motive is that upon which Aristotle relies in the first sentence of the *Metaphysics.*[8] We are confident that if ratiocination can be reconciled with psychological determinism, none of the other less "pure" cases, such as involve means-end selections based upon the combination of "knowledge" and "desire", will present insuperable difficulties. On the other hand, a satisfactory deterministic analysis of motivational ("goal-oriented") behavior, such as the mechanism for selecting instrumental responses tending to find sexual gratification or

avoid social anxiety, might leave us with persistent doubts about the compatibility of determinism with rationality. One could put it succinctly thus: If ratiocination is compatible with determinism, then, so are purposiveness, goal-directedness, motivated choice, contemplation of alternative means-end appropriateness, the "influence of our desires on events", and so forth. Whereas, if determinism is incompatible with ratiocination, Popper's case is proved, whatever might be shown about any or all of these. We therefore submit that the issue will not be prejudiced by our concentrating on the single question, "Is the doctrine of psychological determinism compatible with the existence of human rationality?"

A common ground with Professor Popper can perhaps be found in the following noncontroversial observation: "It is frequently the case that we influence other persons, and find ourselves influenced by them, through the giving of *reasons".* We employ the weak verb "influence", where many deterministic psychologists would prefer to say "determine" or "control", since these stronger words beg the (empirical) question as to precisely how rigidly "clocklike" human behavior is. We do not here enter into the substantive issue, which is qualitatively no different in psychology from that presented by the other biological and social sciences (or, for that matter, as Professor Popper himself emphasizes, the physical sciences) as to whether certain systems are or are not plausibly viewed, on the basis of all available theory and evidence, as ontologically deterministic apart from the question of our "instrumental" ability to predict and control them. As pointed out earlier in this chapter, it is sometimes a rational extrapolation for the scientist to postulate that a system obeys strict laws, i.e., that all the events of a domain are instantiations of nomologicals, even though the limitations of his measuring technology are such that the determination of the initial and boundary conditions of the system make it unfeasible to predict other than statistically. For example, staying away from the human case for the moment, the animal psychologist observes a steady trend toward increased order and regularity in learning curves as he increases his control of the organism's previous history and, especially, his control of the current stimulating field. The smoothness and high-confidence predictability of cumulative response curves obtained in the "Skinner box" is the main reason for its increased use in the study of learning, emotion, psychopharmacology, etc., in preference to the previously popular maze. The lawfulness of the rat's behavior in the Skinner box is sufficiently great (better, for example, than most "physiological" research data) so that a psychologist is sometimes in the position of being able to instruct his research assistant, "See what's the matter with the apparatus", because the curve purportedly produced by a particular rat is one which he can confidently say is *psychologically impossible,* given the animal's previous training and present stimulating conditions. To quote our colleague, Professor Kenneth

MacCorquodale, "These data are impossible; God is a better engineer than Foringer (manufacturer of Skinner-boxes and associated programming and recording apparatus)". It is a matter of degree, not of kind, when the experimental psychologist—or, for that matter, the clinical psychologist, relying upon Freud's investigations—makes the usual scientific extrapolation and assumes until further notice that this fact of increased control (or even well-knit retrospective understanding) suggests that, in the limit, the system would be predictable or, putting it ontologically, that "in itself" the system is deterministic. Again, we do not propose to argue the empirical merits of the substantive questions here (especially the vexed issue, "how do persons differ from rats?").[9]

It is worth noting that there are *non*experimental occasions on which an extremely high degree of predictability, as high as we can normally obtain in an ordinary undergraduate physics laboratory experiment, is present even in the case of complex human behavior involving rational processes. If I take certain rather elementary steps to ascertain that a colleague is in a "normal" state of mind, i.e., that he is not hypnotized or psychotic or drugged or the like, I can predict his answer to certain questions (putting it another way, I can therefore, by asking these questions, control his verbal output) with essentially 100 percent certainty. Thus, we know that, if we present to Professor Popper a certain sort of invalid syllogistic argument and ask him to "comment on this *qua* logician", he will reply, "Illicit distribution of the major", or some synonymous expression. The predictability of this kind of "rational" human behavior is considerably higher than that which obtains in other areas of human behavior (e.g., falling in love, disliking a political figure) and it is also considerably higher than that which obtains in, say, organic medicine, or even in certain domains of the physical sciences (e.g., meteorology).

Furthermore, what appears from the behavioristic standpoint as a practically perfect predictability or controllability of verbal behavior can also be observed introspectively by a nonbehaviorist philosopher of, say, dualist persuasion; and the subjective experience of an individual in this type of situation is consonant with the behaviorist's impression of it, i.e., *one feels subjectively that he is completely powerless to think otherwise.* Of course, he is capable of *speaking* otherwise, and will do so if certain other motivating conditions are provided. For example, one might (as a nefarious Svengali-type psychologist) inform Professor Popper that somebody was going to put a logician's question to him in an effort to "prove the thesis of psychological determinism", and as a result of such instruction Professor Popper might be motivated to show that his behavior is *not* thus neatly predictable, and as an "act of counterwill" refuse to classify the obviously fallacious syllogism as an Illicit distribution of the Major, calling it instead an Undistributed Middle, or saying that it was all right. Here the predictability of what he *would* say is

rather low, but (knowing his position on determinism) the predictability that he *would not* say what we would ordinarily expect him to say as a logician might be very high indeed. But these questions involve his overt speech output. In either case, from the subjective standpoint, he would find himself incapable—we do not say "unwilling", we mean literally *incapable—of* cognizing a fallacious syllogism as being valid. If we, as behavior-engineers, told him, "Sir Karl, we are now going to determine your thought for the next few seconds, provided you are willing to listen to what we say next. Consider the following argument…", he might refuse to listen to us, or decline to read an argument on a sheet of paper (i.e., we might lack adequate control of his orienting and attending behaviors). But *if* he met these conditions, i.e., if he listened to us and thought about what we were telling him, we would have attained substantially perfect control of *what* he would think about what we said. It is worthwhile emphasizing this subjective aspect, because there is a tendency, when philosophers and psychologists quarrel about this matter, to identify *determinism* with *behaviorism.* And while unquestionably these positions show certain historical connections (and temperamental affinities?) they are related by no logical necessity, as Professor Popper has been careful to remind us.

One source of philosophical (and, even more, of one's personal, "existential") rejection of the idea of psychological determinism is our tendency to associate it with those theorists and ideologists that have laid emphasis upon the *irrational* determiners of human thought and action. If you ask the typical cultivated, educated nonpsychologist for his immediate associations to the idea of psychological determinism, he will usually mention Freud, and with fair probability will add Marx, Pavlov, and—depending upon how much he has kept up with the controversy or how recent his formal education—Skinner. Now whereas Freud was a complete psychological determinist and therefore held that the "rational, reality-testing functions of the ego" were determined as much as anything else, he seemed to feel no tension, let alone logical contradiction, between his own very high valuation of rationality in the scientist's thinking, and the notion that such thinking, as much as the scientist's knee-jerk reflex or digestion, was completely determined. The main thrust of Freud's contribution was the *extent* to which irrational forces control "the surface", and the *extent* to which we are often deceived in giving a purportedly rational account of our conduct. Thus one of the Freudian Mechanisms—contributed not by Freud himself but by Ernest Jones—has the title "Rationalization", the process of giving reasons (perhaps even objectively valid reasons) for actions or beliefs that were, in fact, psychologically produced by internal forces of a very different, nonrational character. It was, of course, no *qualitatively* new discovery on Freud's part that people deceive themselves about their own beliefs and conduct; but the

working out of some of the details of the machinery by which this self-deception is carried out, the *quantitative* emphasis upon its being more frequent and more powerful than had generally been supposed, and the elaboration of the *content* of the unconscious processes (e.g., what kinds of impulses are being defended against) have had an impact upon our culture which can hardly be exaggerated. The same is true of Marx who, although he never denied that rational calculation occurred (e.g., when the capitalist asks himself how he can maximize his profits), nevertheless saw a great deal of both individual mental life and cultural development as primarily reflecting economic forces which did not always appear on the surface. Thus we have the stereotype of the kind of vulgar Marxism which would "explain" Darwinism as nothing but the biologist's rationalization of Victorian competitive capitalism, or would "explain" the rise and decline of cubism in terms of the pig iron production of French industry.

Now without entering into the empirical question whether Freud somewhat exaggerated this (admittedly) pervasive influence of the irrational, what we wish to emphasize here is the following: Whether one describes the mind in psychoanalytic terminology, or in the terminology of an experimental psychology of perception and learning, the most deterministic psychologist does not deny the existence of specific cognitive dispositions—of "habits" or "ego-structures"—that are rational in nature. Thus, for example, in Freudian theory we distinguish between the so-called "primary" and "secondary" processes, between the "pleasure principle" and the "reality principle" of mental functioning, between a relatively strong ego—that means, in large part, a realistic or rational one—and a weak ego, as is found in a young child or a regressed psychotic. One should avoid the very common temptation to think immediately, when psychological determinism is under discussion of such determiners as one's unconscious hatred for his father or the subtle influence of a blood chemistry attributable partly to the fried eggs one had for breakfast. Popper himself succumbs to this temptation, in discussing the theoretical predictability of Mozart's or Beethoven's composing, in terms of whether they "had eaten lamb, say, instead of chicken, or drunk tea instead of coffee" (*CC*, p. 11). Sometimes unconscious conflicts or fried eggs are strong enough to impair the ego's rational functions; sometimes, fortunately for the conduct of ordinary affairs as well as the advancement of science, the fried eggs simply produce the necessary biological energy to keep the machine working, but do not, in any psychologically or philosophically important way, determine the *direction* or *content* of the ego's cognitive processes. One should not, in contemplating the social and existential implications of psychological determinism, think only of the fact that a man had a permissive mother or an authoritarian father or a vitamin deficiency, to the neglect of such equally important factors as that he inherited a high concept-

tual intelligence, that he read many books during his teens, that he was exposed to an excellent undergraduate course in logic, and the like.

In this connection it is important to keep in mind the distinction between object language habits and metalanguage habits. Not only do we learn by a complicated mixture of (a) direct reinforcement ("reward") for thinking straight (or, alas, crooked, as the case may be) and (b) by identification with significant figures in our environment who present models of straight or crooked thinking, and (c) by formal precept in school and university, to obey logical rules; we *also* learn a set of powerful metahabits, *such as talking to oneself about the rationality of one's own arguments* (which in the case of philosophically disposed persons may maintain the ascendency in behavior control over many and strong competitive forces). It must be confessed that tendencies of. this sort are not as widespread in humankind as one might desire, and it is a presently unsolved question to what extent this sad fact is a matter of poor education or limitations on basic intelligence and temperament. But even the uncultivated layman of low education does possess a rudimentary set of such metahabits, which can be successfully appealed to, if the counterforces are not too great, to control his behavior along rational lines.

Finally, we all acquire certain *self-concepts* in the process of acculturation. For some persons, the self-concept "I am a reasonable fellow, I do not go around .committing gross fallacies" is as fundamental and important a part of their personality organization as those kinds of self-concepts more familiar in the literature of psychotherapy, such as "I am an unlovable person" or "I am strong, I do not need to depend upon anybody", or "Nobody can tell me what to do!" or "I am a beautiful woman, all men are attracted to me", and so forth. Here again, while the social and clinical psychologists have attempted, with arguable success, to fill in many details about the mental *machinery,* and the family constellation that contribute to mental *content*, it is noteworthy that the basic situation in such matters has always been understood (in its essentials) by thoughtful men. Everyone knows that in discussing controversial matters, say of politics or economics or sexual morality or foreign policy, we often find ourselves trying to judge whether we can successfully appeal to a person's "need to be rational" when that second-order need, involving the possibility of a threat to his self-concept, is, on a particular substantive issue, placed in opposition to what we (from the outside) view as nonrational or irrational commitments. So we may say of a person, "He's a straight thinking, sensible, fair-minded fellow, and you can almost always learn something from him, and teach him something in a discussion; but I must warn you that he admires his father very much, and his father is a classical representative of the old Southern Bourbon type. So on the race question, you have to handle him with kid gloves; there is one

issue where he can sometimes become rather illogical when pressed".

It may be objected that, when the psychologist employs locutions such as "self-concept of being a rational person", he is surreptitiously plugging a nondeterministic concept (i.e., that of *rationality)* into a behavioristic-mechanistic-deterministic theoretical framework in which such meta-categories have no place. One of us (Meehl, 1968) has examined this question at length elsewhere, and the reader may be referred to that[10] for elaboration of the position we adopt on this question. Over against the dualistic, antireductionist philosopher, we hold that there is no contradiction involved in saying, "Jones's thinking is logical [on such and such an occasion]" and saying, "Jones's thinking [including its logicality] is strictly determined by his present neurophysiological state, together with the momentary stimulus inputs; and his current neurophysiological state is in turn completely determined by his antecedents, i.e., genetic equipment interacting with his life experiences".

On the other hand, we must say (against certain kinds of neo-behaviorists) that the psychologist's scientific task—whether its explanatory, predictive, or controlling aspects are emphasized—cannot be carried out *at a molar level of analysis*, unless the psychologist employs certain of the logician's concepts and rubrics in his, the psychologist's, object language. This is not the place to develop that argument in detail, for which development the reader is referred to the article cited above by Meehl. Briefly, the position is that, in order to explain, predict, and control, let us say, the verbal behavior of a logic professor when presented with a formal fallacy such as Illicit Distribution, the psychologist must be able to *characterize the stimulus side*, i.e., the perceptual input to which the logician responds with such an utterance as "Illicit Major". And the point is that the psychologist's use of such concepts as "stimulus equivalence" or "verbal generalizations" must not be employed by him to hide a very important fact: When the psychologist is forced to make explicit the *configural* properties of a stimulus input that will render it "stimulus-equivalent" to the subject logician, so that, for example, the logician can properly classify a syllogism whose terms—except for the logical constants—are terms of which he does not know the meaning, and to which he has never been previously exposed; the only way to characterize this stimulus input is in terms of its *formal structure.* An adequate characterization of that stimulus class, regardless of what terminology the psychologist might employ in describing it, will, of course, turn out upon careful analysis to correspond to the characterization found in a logic text. It is of no substantive interest whether the behaviorist psychologist actually employs the logician's sign vehicle "Illicit Major", since it must be admitted that whatever (nonphilosophical) sign vehicle the psychologist employs, his *definition* of it will involve specifying the very same formal features which the

logician specifies in defining the term "Illicit Major". For this reason we hold that when Skinner[11] speaks of logic being "embraced by our [the radical behaviorist's] analysis", although he is literally correct if he means that reasonable behavior, and the tokening of metalinguistic terms belonging to logic, have a causal history in the learning process; he is incorrect if he means that this behaviorist "embracing" involves the liquidation, or a showing of irrelevancy, of the logician's *categories* in a psychological analysis of "reasonable verbal behavior". *The molar behaviorist psychologist who concerns himself with language and with rational, cognitive, ego-functions must reconcile himself to the fact that he cannot dispense with the logician's formal categories.*

We mean by this something much stronger than what would be meant if we said it about a physicist or a botanist. Every scientist has to come to terms with the logician in two ways, namely, (a) He must exemplify logical processes in his object-linguistic discourse, i.e., he must think rationally about his subject matter; and (b) Since a great deal of scientific discussion and scholarly writing is not simply reports of observations, or formulation of theory, but is critical discussion of theories (their relationship to one another and to observations, the validity of one's own and other people's inferences, and the like), the scientist must also make use of the logician's categories in his metalinguistic discourse, i.e., in the process of scientific criticism. In respects (a)–(b) psychology is not essentially different from the other sciences. However, there are no other sciences in Comte's pyramid of the sciences below the level of psychology (we are assuming economics, sociology, anthropology and political science all appear "above" psychology in this well-known pyramid) which are forced to employ the categories of the logician or philosopher in their object language.

It is not clear why this necessity should be distressing to a biological or social scientist, but for some reason it often seems to be. Whether it should distress the philosopher depends upon whether there is some kind of paradox or contradiction involved in *metalinguistic* terms, such as "valid", or "Illicit Distribution", appearing in the *object* language discourse of an empirical science. But, if this represents a problem for the philosopher, it represents a technical problem in logical theory, so we content ourselves with merely calling attention to the oddity. (Our colleague, Professor Keith Gunderson suggests that, since this unavoidable reference by the molar behaviorist to formal categories is so evident, any logician's theory about object language/metalanguage relationships that precludes it is, *ipso facto,* suspect.)

A final question concerning any reductionist-determinist view of the psychology of beliefs, arguments, inquiry, criticism, and the like is the question, "Would a complete causal account, *formulated in terms of the microlevel* (e.g., electrical and chemical events at neural synapses) be incompatible with

our intuitive conviction that our beliefs and actions are influenced by reasons?" It is our contention, as opposed to Sir Karl's, that there is no such incompatibility, although at first impression it does appear that there must be. Suppose one takes the expression "a valid reason" as designating a kind of abstract Platonic universal which *in some sense* "exists", and would exist even if there were no thinking brains (a position we are not here espousing, but will adopt as a premise *arguendo,* since it is the one most unfavorable to the determinist analysis of human thought and action); then we hold that "the existence of a valid reason" (in some such abstract, metaphysical, Platonic sense) is a question belonging to the critical domain of logic, broadly conceived, But the *thinking* of such a reason by a living, concrete human person, the *stating* of a reason, the *hearing* of a reason, the *mentioning* of a reason, *the tokening of a sentence which expresses a proposition which is a good reason for another proposition*—these are all events very much "in the world" and "belonging to the causal order". And we maintain that one does not have to conflate reasons with causes, or to commit the fallacy of psychologism in logic (or the naturalistic fallacy in ethics), to be justified in saying that, although the validity of an argument, or the soundness and cogency of a reason, is an abstract Platonic truth about universals, nevertheless the thinking of a reason is an event, is a something which happens in space-time, in a living brain (just as the uttering of a valid reason happens in a human larynx), and instantiates nomologicals.

Admittedly there is something initially strange about the notion that a man's beliefs or actions are influenced by reasons, i.e., as we say a man is to this extent and in this matter "reasonable", and nevertheless hypothesizing—as a promissory note, until further notice, as an orienting "faith" of the scientific investigator—that the brain processes involved could, in principle, be formulated by Omniscient Jones *at a level of causal analysis which would dispense with the logician's categories.* But, although this does seem to us a bit odd, and to some readers will doubtless be highly paradoxical, we are not persuaded by Professor Popper's paper, or by any arguments that have been thus far brought to our attention, that there is anything contradictory in it. What must be understood, if we are right in our view of the matter, is that when one gives a complete causal account of a physical process at a certain level of analysis, *he does not thereby claim,* in making a metaclaim to "causal completeness", *that he has asserted everything true that could properly be asserted about the system.* In other words, it is the difference between telling the truth and telling the *whole* truth.

It is important to notice here that, whereas Professor Popper's distinction between "Compton's problem" (puzzles about determinism) and "Descartes's problem" (puzzles about the mind-body relationship) is a valid and useful one, at this stage of our discussion they are seen to be intimately

related; and if the reader will turn back to the series of quotations from Popper at the beginning of this section, he will notice that Professor Popper himself has at times conflated them, we think perhaps unintentionally but unavoidably. Although the (Compton) question "How can I be rational and purposive if determined?" can be seen as presenting a philosophical problem even within a framework of radical (ontological) mind-body dualism, its bite is greatly sharpened if, instead, the determining (and determined) events are all set in a purely physicalistic ontology. Roughly put, it is bad enough if my "mind" is determined; but it is worse if my "mind" is nothing but a complex configuration of events or states, a sequence of occurrents that determine-istically befall continuants that are themselves "nonmental" in nature. This fusion of the Compton problem and the Descartes problem has, of course, a venerable history, being already formulated clearly and powerfully by Plato (in the *Phaedo,* 98c – 99b) and Aristotle (in the *Metaphysics,* Book 9, Chap. 6, 1048b—18-30). It has received considerable attention from several contemporary Oxford philosophers, and from the Americans A. I. Melden and Richard Taylor, to mention only notable examples. It is perhaps foolhardy for us in this limited space to attempt a resolution of such an ancient and recondite question, to which so many able thinkers have addressed themselves (and, on the current scene, emerged with answers very different from ours). But we are satisfied that no one has formulated the determinist-monist analysis of rationality (or purposive action) *quite* in the manner we propose, and that our kind of "levels" picture of the situation, whether ultimately refutable or not, will at least constitute a contribution to the ongoing philosophical discussion of this venerable controversy.

Consider first a nonpsychological example. If I give a detailed, blow-by-blow mechanical account of a sequence of operations carried on by an ordinary desk calculator, I can properly say, in conclusion, that "causally speaking, nothing has been 'left out' of this account". But it is evident that such a detailed account in terms of the laws of mechanics (as applied to the structure-dependent properties of the machine in respect to its inner workings) does not *have* to contain such (true) statements as, "The machine is dividing 3,746 by 125". Yet that is, quite literally, what the machine is doing. It does not occur to us, in such simple, inanimate system cases, to postulate the existence of some kind of an "arithmetical" *deus ex machina* as a necessary addition to the causal system first described. Nevertheless, one can shift the level of description upward to a more "molar" level, as an instructor in a statistics laboratory would do in talking to a student about how to operate the machine. At that (more "molar"-behaviorist) level of analysis, the instructor formulates quasi-causal laws (they are tight nomologicals, as long as the machine is not broken or worn out) in which such concepts as addition, division, multiplication and the like occur in the formulation.

We think that there is a temptation, when philosophers consider the implications of psychological determinism for the possibility of human rationality and genuine criticism, to move from the (correct) statement, At level L_c of causal analysis, not everything which might truly be said has been said" to the false (as we think) statement, "Therefore, at level L_c of analysis, the causal account is radically incomplete". If certain definitions are given at one level of analysis, and then certain reductions are carried out (given acceptance of a suitable theoretical network), it will then be clear that the things which were left unsaid follow necessarily from those things which were said, when the latter are taken together with the explicit definitions. So that, whether or not it is literally correct to say of a microphysiological account of a complicated human thought process that the account "leaves something out", depends upon a clarification of what is meant. If the speaker means to say that something has been left unsaid which could truly be said, he is right; but if he means that the causal account will not be complete unless some additional *theoretical entity* is introduced into the causal chain, then—assuming that the reductionist thesis is empirically correct—he is wrong.

In speaking of "reduction" and "definition" in the preceding paragraph, we are perhaps inviting a major misunderstanding which we shall now endeavor to forestall. Our position is not, we trust, a surreptitious reduction of the *logical* to the *physical,* as if to say that conceptual relations (e.g., class-inclusion, formal deducibility, contradiction) could be defined in terms of "nonlogical" notions (e.g., mass, rigidity, proximity, protoplasm, synaptic resistance). We take it that everyone—whether philosopher, physicist, psychologist, computer engineer, and whether determinist or not, "physicalist" or not—would agree with Sir Karl that any such reduction is impossible in principle. Logical and arithmetical relationships are *sui generis,* not definable in terms of physical or biological categories or dimensions, and we wish to make it crystal clear that we accept this truth unqualifiedly and without equivocation.

In what sense, then, can we properly speak of "explicit definition" in the preceding line of argument? Reflect again upon the desk calculator problem. Replying to an imagined critic (one who has Cartesian, antibehaviorist, anti-physicalist or emergentist views about these machines), and who complains that our detailed micro account of the machine's transitions "leaves out the main point, namely, that the calculator is adding the numbers 4 and 3 to get the (valid) answer 7", we say the following: "It is a true statement that you offer as supplementary to our mechanical account; the machine is, quite literally, 'adding numbers', and, furthermore, it is 'getting the correct answer'. And it is true that our mechanical description did not anywhere include this arithmetical assertion. You are therefore entirely right in saying that our account does not 'say everything that can properly be said'. What we

deny is that we have 'left something out' in the causal sense; specifically, we deny that there occurs any event, state, or process involving any *theoretical entity* [= entity playing an explanatory role in the nomological net, possessing causal efficacy] over and above the physical entities included in our 'nonarithmetical' account". We wish to maintain that the following statements are jointly compatible:

1. Our causal account A_c of "how the calculator works", as a physical mechanism, is complete.
2. The words "sum", "addition", "integer", etc., do not occur in A_c. (Names *of* integers may, of course, occur in the account A_c; mathematics is part of the object language of mechanics, of course. We do not talk metamathematics in describing the machine's physical operations; we do not even *mention* arithmetical theorems. We do, however, *rely upon* such theorems; and, of course, we use numerical *concepts* [e.g., "the second gear underwent three forward tooth-displacements"] .)
3. "Number" is not a theoretical (causal) entity contained in, or acting upon, the machine.
4. The machine (literally) *adds numbers.*

We invite the reader to try developing a genuine contradiction from these four statements. We think there is none. What there is, is a kind of oddity—the kind of oddity that wears off, however, with sufficient familiarity. Not that we would for a moment countenance invoking oddity as a negative criterion, especially in these deep matters.

On the positive side, perhaps the shortest formulation (aimed at therapeutic erosion of the oddity-response) would be something like the following: Abstract "entities", such as formal (numerical, logical, set-theoretical) relations considered as Platonic universals, are not "in" the machine, in the sense that its *parts* are *in* it. Nor are they "in" the *event*-sequence, as, say, a wheel displacement is. But the physical entities (whether continuants or occurrents) which *are*, literally, *in* the machine, do (in some respects) exemplify those abstract universals. The cardinal number 3 is not, obviously, in the machine. But sets of structures and events of cardinality 3 are there.

In discussing with colleagues our approach to this aspect of Popper's problem, we were once met with the objection that our analysis really consisted of saying that the desk calculator (or a "logic machine", or the human brain) is a physical model satisfying the laws of arithmetic, which *all* physical entities necessarily satisfy; from which, the critic argued, it would follow that no calculators (or brains) can err. This complaint rests upon a confusion among physical levels. We must be clear about *which* physical entities and

processes are taken as elements and relations of the physical model corresponding to the elements and relations of the calculus. It is, of course, true that the physical set formed by conjunction of a set of three iron atoms and a set of two iron atoms is a set having cardinality 5, even though the gear in which these five atoms lie is part of a worn-out calculator (which "makes arithmetical mistakes"). But the cardinal number of interest is not that of the constituent atoms, it is the number of tooth-displacements. Or, treating the machine "molar-behavioristically", the cardinal number of interest is either (a) The number of punch-and-cumulate operations or (b) The cardinal number designated in English, by the numeral [= "3"] on a key punched and cumulated once. And it is obvious that *these* (= "molar") physical events do not necessarily satisfy the laws of arithmetic. *If* the wheels are not worn out, prone to slippage, etc., then three punchings and cumulations of key labeled "1" will result in a state of three-tooth-displacement; this state will persist until two further unit-additions occur. It being a theorem of arithmetic that 3 + 2 = 5, we know that 3 physical displacements "plus" 2 physical displacements leads to a terminal state of 5 net displacements. It is (trivially) analytic to say that "If the machine satisfies the axioms, it satisfies the theorems". But the point is that, whereas its constituent atoms satisfy the axioms, the "mo1ar"-level parts and processes may fail to be a model of arithmetic, when the physical operations occur in time and the numbers are used to characterize resulting positions rather than historical facts. The statements S_1: "Gear G has undergone three forward displacement events during interval $(t_1 - t_0)$" and S_2: "Gear G has undergone two forward displacement events during interval $(t_2 - t_l)$" jointly entails S_3: "Gear G has undergone five forward displacement events during interval $(t_2 - t_0)$". This truth of arithmetic, that $S_1.S_2 \rightarrow S_3$, cannot fail to be satisfied by the machine, worn-out or not. But the conjunction $S_1.S_2$ does *not* entail S_4: "At t_2, Gear G is in a state displaced five steps from its state at t_0'. Gear G may have "slipped" backward at some time during the interval. And similar considerations apply a fortiori when a causal interaction between different gears is supposed to model arithmetical operations.

Unavoidably, any causal reconstruction of rational mental processes in terms of brain events will suffer, at the present state of our knowledge, either from extreme generality—amounting to little more than a restatement in pseudo-brain-language of such general formal concepts as "model"—or, if it becomes more specific than this and attempts to deliver the goods, so to speak, *scientifically,* it will be on the fringe of current empirical knowledge and readily objected to by the critic as not only unproved but excessively speculative. We do not take it as our task, in examining Professor Popper;s objections, to present a brain model. And since his objection is essentially philosophical (rather than directed at what is wrong with any specific sub-

stantive theory of learning, perception, or thinking in the present state of the behavior sciences) we are confident he will not complain in either of these two ways, so long as we make quite clear which enterprise we are engaged in, i.e., a highly generalized statement of the conditions for the brain to think rationally although determined, or, on the other hand, a specific physicalistic *example* which exhibits the philosophical point we wish to make but lays no claim at all to scientific correctness.

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 truly 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 microaccount 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 behavioralistically they were originally acquired by social conditioning, considering them as now playing the role of Sellars's *mental word*[12]*—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 enormous complexity of adequately explaining the words "correspond" and "designate" in the immediately preceding text.)

A physically-omniscient neurophysiologist [=Omniscient Jones estopped from metatalk about logic] can, we assume, identify these four brain events b_M, b_P, b_S, b_C on the basis of their respective conjunction 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 assembly, 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) microstructural statements including microstructural statements about functional connections. We take it for granted that "learned functional connections" *must* be embodied in microstructure (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 subclasses 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 $(L_{phys} \cdot S_t)$, 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 metaconcept 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 away, 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, whereas 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 Gödel to notice, *qua* logician, that these brain events constitute a physical model of a subcalculus 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 metatalk 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_i b_j b_k) (b_j b_k b_l)…*embodies* the syllogistic rules. Their defined physical structure plus the physical laws of brains 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 meters, of Mount Everest—is literally a condition for the inference to occur.

Comes now the emergentist or Cartesian dualist advancing an objection as follows: "Even granting, which I do not, that there are any such strictly determimstic brain nomologicals as L_{phys}, and assuming *arguendo* that your concept of a physically-omniscient neurophysiologist (who is ignorant of

metastatements in logic) could actually carry out the derivation of T_b from $(L_{phys} \cdot S_t)$, I must interpose a philosophical objection which is surely available to us upon present knowledge, and does not rely upon speculations, whether optimistic or pessimistic, about the future development of brain science. My philosophical objection is a very simple one, and it is this. While you, in your imaginary reconstruction of the derivation possibility for the physically-omniscient neurophysiologist, can carefully avoid any reference to the syllogism, or to the formal validity of the inference which is 'embodied' in these brain events, *you yourself* do have in mind that this is the brain of a thinking human being, and you do intend me to understand that these brain events are his valid thinkings. Yet you say that you give an exhaustive account of what happens, and that you can explain, i.e., derive nomologically, given the structural statements S_t, all that went on. Now if it is true that you can derive everything that goes on within the object language of the physically-omniscient neurophysiologist, then the fact that this brain event sequence is [known to you and me but not to him] a process of valid inference, that the individual whose brain we are studying is *correctly thinking a syllogism in Barbara,* turns out to be an irrelevant fact, a supernumerary, something that doesn't make any difference at all. Whereas I want to hold that this is the most important fact about these happenings. How can you say that you have given a complete account of what is taking place in this person, that you have 'explained everything' about the happenings under consideration, when you have not anywhere said the most important thing there is to say, namely, *that he is making a valid inference?* If the physically-omniscient neurophysiologist can derive the brain theorem T_b without even knowing that inference is going on in this person's brain (I should of course prefer to say 'mind', but I will make concessions to your strange materialist verbal habits), then the validity of the inference makes no difference. Whereas, of course, I know perfectly well, brain physiology aside, that whether a certain inference is permissible makes all the difference in the world, and you yourself have said the same thing earlier in the present discourse. I simply do not understand how you can say that the account is causally complete when it leaves out the most important fact, which is a *logical* fact, namely, that a valid inference is being made when the person passes from tokening 'M is P' and 'S is M' to tokening 'S is P'."

Now the first thing to see about this Cartesian or emergentist objection is that it contains an important element of truth, which it is both dishonest and unnecessary for the determinist-materialist to deny. The objector points out that, in the microphysiological account of the brain events which is assumed to be offered by the physically-omniscient neurophysiologist, something that is literally true has not been said, to wit, *that a valid syllogism is being tokened.* Having freely admitted this, our next question is, does it

follow from the fact that something which could be truly literally said has not been said, that the causal account is incomplete? This is the crucial issue, and we want to urge that the proper answer to this question is in the negative. But, of course, in order to examine this question, one has to arrive at some suitable convention for the use of the metalinguistic expressions "everything that can be said" and "a complete causal account of the events".

Let us first point out that there are many obvious and noncontroversial cases, lying quite outside the realm of the mind-body problem (or the determinism problem, or the intentionality problem) where it is clear that a complete causal account of a series of events can be given without explicitly saying everything that would, if said, be literally true. Examples are so numerous as to be almost pointless, so we will mention only one. Suppose that I give a causal account of the displacement of the piston in a cylinder by an expanding gas, and this account is given at the microlevel—in the extreme case, by the long, but finite, set of sentences describing the components of momenta, the positions, the impacts on the wall of the container, etc., of the individual gas molecules. Nothing is left out of this account from the standpoint of physical understanding. I do not, however, mention that the number of molecules in this batch of gas is prime. Let it not be objected that this is a silly instance. That the number of these molecules is prime is just as much a physical fact as is the set of numbers characterizing the components of their respective momenta. It is not, however, a fact we find it necessary to *mention* in giving a complete causal account of how the gas displaces the piston. This example suffices at least to show, what is really a rather trivial and unexciting truth, that one can say everything that is causally relevant in accounting (at all levels of causal analysis) for a certain phenomenon, without saying everything that is literally true about it.

But we will readily concede to the Cartesian critic that this example is, while sufficient to prove the just stated general thesis, not a fair analogy to the brain theorem problem. Because what he is objecting to in our brain theorem case is not only that we have left out something that is literally true (that the brain is thinking about a syllogism in Barbara), but that the *outcome* of the thinking is critically influenced by the *logical fact* that the inference is valid. And while it is literally true that the number of molecules in the gas is prime, this is not a fact which makes any difference to the outcome.

This last locution, the phrase "makes any difference to the outcome", contains, we think, the core of the objection, and also provides us with the essence of our answer to it. What do we mean when we say that something's being the case, or not the case, "makes any difference"? Depending upon one's views concerning causality, a number of things might be meant; but we assume that at least a minimum condition for its being appropriate to say that "such and such makes no difference" is that if such and such *had not* ob-

tained, the outcome *would have been* the same as the outcome in fact was when such and such *did* obtain. Whatever else one might mean about something "making no difference", he surely must mean at least this counterfactual. Now is it literally correct to say, as the critic does here, that the causal account provided by the physically-omniscient neurophysiologist shows that the validity of the inference, as a formal truth, makes no difference? That is, does it follow, from the metafact that the physical microaccount provided by the neurophysiologist does not contain any reference to the formal structure (revealed by the subscripts on the *b*'s), that *if this formal structure were not present, the terminating brain event* (b_S b_C b_P*) would occur nevertheless?* Clearly not. The brain theorem is stated in terms of *b*'s with distinguishable subscripts, and these subscripts are formally related, *within the theorem* T_b, in a certain way. Thus, we do not have a brain theorem stating the following: "Whenever brain events (b_S b_C b_M) and (b_P b_C b_M) occur in close temporal contiguity, they are immediately followed by brain event (b_S b_C b_P)". In fact, if the brain under study is that of an always-rational man, we will instead have a countertheorem here, which states that this particular contiguity will never be immediately followed by the third event, i.e., this being the brain of a perfectly rational man, he never commits the fallacy of Undistributed Middle. Now the point is that the physically-omniscient neurophysiologist *does not have to mention* this formal feature about the *b*-subscripts in order for his account to be causally complete, because the statements he *does* mention jointly entail, as a matter of nomological necessity, that the brain events will be related in such and such a way, i.e., that they will in fact be models of the syllogism in Barbara, a consequence which is revealed by the syntactical configuration of the subscripts. And it is literally false to say that the same "conclusion" (b_S b_C b_P) would be tokened in this brain if the subscripts designating the physical events preceding this conclusion-tokening had been different from what they in fact were. In a way, the point is really quite trivial, being essentially no more than to say that if any set of properties makes a causal difference, then any other set which can be explicitly defined in terms of the first set also "makes a difference", for if those higher-level properties were not in fact what they are on a particular occasion, then, since they are defined explicitly in terms of the lower-level properties, it follows that the lower-level properties could not have been what they in fact are; and therefore the outcome would have been physically different.

It still seems admittedly a little strange that one does not have to mention the validity of the syllogism thus exemplified or embodied by the brain processes in giving a complete physical-causal account. We believe that this strangeness arises mainly from the fact that the causal account offered by the physically-omniscient neurophysiologist *begins* with certain microstructural

premises contained in the collection of sentences S_t and that *these sentences collectively constitute the physical description of the microstates which embody the dispositions to compute rationally in this brain,* dispositions that have been socially learned when the individual acquired the habits of "thinking straight". That is the historical account of why these microstructural facts are what they are. But we may start our explanation of the brain theorem T_b at a given moment in time, forgetting about the learning history; then we represent only the *results* of social learning (to think straight) by statements about microstructures at the synapse, considered neuron by neuron. This way of describing the matter of course yields a paradoxical effect, because it now appears that the all-important "fact of rationality" has been left out of the picture. But the point is that the all-important fact of rationality has not really been left out of the picture, *it has instead been stated* (admittedly opaquely) *by referring to the microstructures which embody the "rational dispositions".* There is, of course, a familiar sense in which one may be considered to have expressed everything in a theoretical system when he expresses the postulates, in that he is considered to "assert" implicitly all of the theorems that follow therefrom. From this point of view, our physically-omniscient neurophysiologist therefore "asserts" the brain theorem T_b, and he therefore "asserts" the syntactical relationships holding among the b_i-subscripts in the terms that are contained in the brain theorem. And he therefore "asserts" although not expressly (since he speaks no metalanguage), the crucial fact that this brain is a rationally reasoning brain.

Whether a purportedly complete causal account deserves criticism because of its failure to assert explicitly one or more of the necessary consequences of what it *does* assert, depends upon several considerations such as human social interest, level of analysis as defined by the pragmatic context of one's investigation, and the like. Therefore to leave out a statement that follows from the conjunction of other statements, to the effect that the number of molecules in a contained gas is prime, would not ordinarily strike us as a very serious omission, although it would be a valid comment that "a true statement was left unsaid". But even if we remain at the level of inorganic processes, where purpose and knowledge are not attributable to the system under study, it is easy to think of instances in which something is left unsaid that one would normally think of as a very serious omission, yet that something is not something that one *must* say in giving a complete causal account of what took place. Example: I drop an ice cube in a cup of boiling tea, and physically-Omniscient Jones gives a detailed microaccount of the ensuing process of its dissolution. That is, he considers the tea, molecule by molecule, and considers the face of the ice, molecule by molecule, describing each molecular collision, the interplay between the mechanical forces of impact and the intermolecular forces tending to maintain the solid state of the ice

crystal, etc. We end up with all of the water molecules that were in the ice being dispersed among the water molecules in the tea, and our story is closed. Here is a situation where, unlike the silly case of the prime numbered gas molecules, we are strongly impelled to say, at the close of the account, "Well, that's all very fine, but you know you never said once that the main thing going on here was that hot tea was losing heat to cold ice and cold ice was taking up heat from hot tea, and the result of this whole process is that an ice cube in a cup of hot tea has dissolved and disappeared. You didn't say that. You left that out. *And what you left out is the most important thing that went on".* Our point is that the question of "importance" refers to a number of considerations, which collectively may or may not lead us to be critical of an account which omits to mention something that is literally true and that in fact follows deductively—given certain already available nomologicals and explicit definitions—from the statements that were made in the account.

In the case of the human brain thinking syllogistically, it is obvious that there are at least two kinds of things that could be said that are not said in stating and deriving the "brain theorem" T_b, and which make us critical of the account. The first is the role of "purpose" (motivation, intention, goal-orientation) in the thought processes, which we might express either with reference to the reinforcement history of this individual (as a consequence of which his brain cells have acquired the syllogistic dispositions that they have); or, alternatively, by reference to the state of distress (surprise, disconfirmed expectancy, shame or guilt, or whatever) that the individual is disposed to experience if he finds himself committing a formal fallacy. This motivational feature is one that has not until recently engaged the attention of computer-simulation investigators, and it will unquestionably come in for a great deal of systematic attention in the near future. See, for example, Simon.[13] It was, of course, not our intention to pretend that such goal-oriented factors are absent in the above account, but we presume that Sir Karl will agree that, in spite of their intimate relations to human thought, *intentionality* and *intention* are different concepts; that one thinks rationally and that one desires to think rationally are two different facts, although they are intimately related.

Secondly, we also omitted reference to those metatalk dispositions of the sort that we imagined our physically-omniscient neurophysiologist to lack. We restricted his omniscience to physical events, and kept him out of the metalanguage of logic for our purpose in discussing the physical derivability of the "brain theorem". But it does not appear to us that any new issue of principle concerning determinism and rationality is raised by the introduction of the metalanguage. Whether there are any insuperable set-theoretical or Gödel-related difficulties in computer-hardware analogs to the human mind is not the province of this paper and is beyond our scholarly competence. Fortunately for us Sir Karl has not made such considerations the focus of his

attention (except insofar as there is a set-theoretical problem involved in the world-calendar idea, dealt with in Section I above).

Of course, in the real world no such "brain theorem" as T_b will be validly deducible from realized conditions S_t together with the cerebral nomologicals L_{phys}, inasmuch as all actual human thinkers (even philosophers!) do at times commit formal fallacies. It does not appear to us that this qualification is germane to the analysis we are proposing, however; it merely indicates that the human brain is, as we readily agree with Sir Karl, not a complete "clock" but a mechanism having some "cloudy" properties. For expository purposes we have considered the idealized case, that is, the brain of a thinker who never reasons fallaciously. The necessary modifications are easy and obvious ones which do not, as we think, affect the point of our line of argument. There are two ways to modify the above brain theorem example which would render it more realistic, one of which accepts a determinist picture of the world and the other a quasi-determinist picture with "random" components included as part of the system. On a completely determinist view, the conjunction $S_t\, L_{phys}$ is explained in such a way that the various causal *sources* of fallacious reasoning, e.g., of failure to conclude validly in Barbara, are ruled out as initial conditions by restrictively characterizing S_t in sufficient microdetail. (Another way of doing this, still preserving the determinist position, is to consider three classes of determiners, one of which is the structural determiners, the second of which is again L_{phys}, but the third of which is "momentary state-variables" such as emotion, or the presence of some kind of unconscious Freudian dynamism, or unusual stimulus input or the like.) Alternatively, however, we may choose to do justice to what we, like Popper, suppose to be a certain cloudlike feature in the cerebral quasi-clock (even as causally understood by Omniscient Jones) in which case we must replace the brain theorem T_b in its nomological form by a brain theorem T_b' in stochastic form. In this case, the relationship between the stochastic brain theorem T_b' and the formal structure of the syllogism in Barbara (as shown forth in the subscripts on notation designating the brain states about which the theorem speaks) would be analogous to the familiar distinction between the statistical data that constitute an anthropologist's or a psycholinguist's "descriptive semantics" of a language, and the idealized semantical rules (*prescriptive* sense of "rule") which the dictionary writer, semantician, or logician sets down on the basis of this descriptive survey. Obviously, the overwhelming preponderance of successor states to the "premise-tokening" brain states will, in an intelligent non psychotic brain, be such as to reveal the same formal structure as in our idealized example. So that a perspicacious logician would have no more difficulty discerning the formal structure in the stochastically formulated theorem than he would in the nomological form, since he would immediately notice the high probability outcome ($b_S\ b_C\ b_P$) and then discern that its

subscripts are Barbara-related to the subscripts on the antecedent cerebral events. But, of course, while identifying the usual (valid, "rational") sequence would take care of the needs of the logician who is attempting to understand whether, and how, this particular brain "thinks correctly"; the molar psychologist has an equally great interest—whether Freudian or less clinical in bent—in understanding causally why the brain theorem is merely stochastic, and in identifying the major classes of interfering factors which operate to produce departures from "validity" on certain occasions.

CONCLUSION

We must leave to the reader to judge, in the light of Sir Karl's rejoinder at the end of this volume, as to whether our criticisms of his argument are sound or not. In either case, we anticipate that his reply, like the papers to which we have reacted herein, will be stimulating and illuminating. We have not taken the reader's time nor the volume's limited space to emphasize our numerous points of agreement with Sir Karl (as, e.g., the untenability of those muddled and malignant varieties of societal determinism he so brilliantly refutes in *The Poverty of Historicism* and *The Open Society and Its Enemies)*. Nor have we bestowed superfluous encomiums upon this great man. We have been engaged in an attempt to control his, and the reader's, cognitive processes—but by the method Sir Karl has so well defended in all of his writings, the method of rational criticism. We are confident that his own position will grow in clarity and depth by reaction to these criticisms of ours, as we have been moved (quasi-determined) to think afresh about the determinist position as a result of his objections.

It is perhaps unnecessary to add some final disclaimers as to what we have *not* been doing in the preceding essay. We have not been trying to make the positive case for radical determinism, which neither of us can presently hold for physics, and which is a methodological prescription as well as a promissory note, for molar psychology—the latter unlikely, for reasons Sir Karl persuasively adduces, ever to be fully paid. We have not been engaged in making a positive case for materialist monism (the "identity thesis" as a solution of the ontological mind-body problem), regarding which we both entertain doubts, mainly because of semantic difficulties with sentience (raw feels) but also, in part, due to the puzzling data on telepathy and precognition. We have not attempted to reduce the categories of logic or semantics to those of physics or biology, knowing as we do that all such efforts necessarily involve philosophical mistakes of the most fundamental kind. Finally, we have not been so foolish as to offer a specific brain model, or to propound a molar theory of cognition, which is one of the most primitive and least understood

domains of scientific psychology. Our purposes have been much more modest and limited in aim, namely, to criticize the philosophical arguments by which Sir Karl has tried to show—thus far unsuccessfully, as we think—that a doctrine of determinism or quasi-determinism (= physiological determinism qualified by cerebral randomizing systems) is *incompatible* with the admitted facts of (a) practical unpredictability, (b) creative novelty, and (c) rationality (insofar as it exists) in human affairs. If our critique contributes constructively to the development of Sir Karl's thought and to further clarification of the profound and difficult issues involved, we shall be satisfied.

HERBERT FEIGL AND
PAUL E. MEEHL

MINNESOTA CENTER FOR PHILOSOPHY OF SCIENCE
UNIVERSITY OF MINNESOTA
SEPTEMBER, 1968

NOTES

[1] K. R. Popper, *Of Clouds and Clocks: An Approach to the Problem of Rationality and the Freedom of Man* (St. Louis, Missouri: Washington University, 1966.)
Also: "Indeterminism in Quantum Physics and in Classical Physics, I and II", *The British Journal for the Philosophy of Science*, **1**, No. 2 (1950), 117-33, and No. 3, 173-95.
Conjectures and Refutations: The Growth of Scientific Knowledge (London: Routledge & Kegan Paul, 1963; New York: Basic Books, 1963), esp., "On the Status of Science and of Metaphysics", pp. 184-200; "Language and the Body-Mind Problem", pp. 293-98; "A Note on the Body-Mind Problem", pp. 299-303.

[2] They are not decidable for the simple reason that their formulation requires universal as well as existential quantifiers. (But many scientific propositions have this form.)

[3] K. R. Popper, "Indeterminism in Quantum Physics and in Classical Physics", Part I, *The British Journal for the Philosophy of Science,* **1**, No. 2 (August, 1950), 117-33; *Ibid.,* Part II, **1**, No. 3 (November, 1950), 173-95.

[4] Although it is worth mentioning that the very clocklike electronic computer, programmed to discover proofs in the elementary propositional calculus by heuristic means (i.e., without the use of a systematic algorithm or decision procedure) succeeded in finding a proof for Theorem *2.85 of *Principia Mathematica* which required three steps and relied upon five earlier theorems, whereas Russell and Whitehead's proof required nine steps and relied upon seven additional theorems, including a lemma whose only use in *PM* is to help prove Theorem *2.85 (and which in turn required eleven steps, yielding a total of twenty steps needed by these human logicians). This strikes us as a rather impressive display of rational creativity by the clocklike computer, and those with expertise in the field assure us that they have hardly begun to tap the machine's resources. Prophecy by two nonexperts: A next major advance might well be metaprogramming for a wide class of mathematical structures, with a randomizer, so that the machine will *invent theories.* It will still have its built-in plus programmed limitations—there will be "theories which it cannot think". But then, as Kant would emphasize, the same is true of us. See A. Newell and H. A. Simon, "The Logic Theory Machine: a Complex Information Processing System", *Transactions on Information Theory* (Institute of Radio Engineers, September, 1956), Vol. IT-2, No. 3, pp. 61-79; A. Newell, J. C. Shaw, and H. A. Simon, "Empirical Explorations of the Logic Theory Machine", *Proceedings of the Western Joint Computer Conference* (Institute of Radio Engineers, February, 1957), pp. 218-30; A. Newell and J. C. Shaw, "Programming the Logic Theory Machine", *ibid.,* pp. 230-40; A. Newell, J. C. Shaw, and H. A. Simon, "Elements of a

Theory of Human Problem Solving", *Psychological Review,* **65** (1958), 151-66; L. T. Allen Newell, J. C. Shaw, and H. A. Simon, "Note: Improvement of a Proof of a Theorem In the Elementary Propositional Calculus" (unpublished Carnegie Institute Proceedings Working Paper No.8, Carnegie-Mellon University, January 13, 1958).

[5] K. F. Schaffner, "Approaches to Reduction", *Philosophy of Science,* **34**, No. 2 (1967), 137-47.
John Kekes, "Physicalism, the Identity Theory and the Doctrine of Emergence", *Philosophy of Science,* **33**, No. 4 (1966), 360-75.
C. G. Hempel, *Aspects of Scientific Explanation* (New York: The Free Press; London: Collier-Macmillan, 1965), pp. 259-64.
J. J. C. Smart, *Philosophy and Scientific Realism* (New York: Humanities Press, 1963). Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace & World, 1961), pp. 367-80, 433-35.
Gustav Bergmann, *Philosophy of Science* (Madison: University of Wisconsin Press, 1957), pp. 140-71.

[6] K. R. Popper, *Of Clouds and Clocks: An Approach to the Problem of Rationality and the Freedom of Man* (St. Louis: Washington University, 1966). Hereinafter cited as *CC*.

[7] For a remarkably sophisticated treatment of this issue, cf. S. Freud, *The Unconscious* (1915, Standard Edition, XIV, pp. 161-215; [London: Horgarth Press, 1957]). The *locus classicus* for a molar-behavioristic defense of "cognitions" as inferred theoretical entities in infrahuman animals (and, hence, not known to have a subjective, phenomenal aspect) is, of course, Edward Chace Tolman's great *Purposive Behavior in Animals and Men* (New York: Century Company, 1932).

[8] "All men by nature desire to know".

[9] The interested reader is referred to such sources as W. Honig, ed., *Operant Behavior: Areas of Research and Application* (New York: Appleton-Century-Crofts, 1966); L. P. Ullmann and L. Krasner, eds., *Case Studies in Behavior Modification* (New York: Holt, Rinehart & Winston, 1965); L. Krasner and L. P. Ullmann, eds., *Research in Behavior Modification* (New York: Holt, Rinehart & Winston, 1965); R. Ulrich, T. Stachnik, and J. Mabry, eds., *Control of Human Behavior* (Glenview, Ill.: Scott, Foresman, 1966).

[10] Paul E. Meehl, "Psychological Determinism and Human Rationality: a Psychologist's Reactions to Professor Karl Popper's *'Of Clouds and Clocks'* ", *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1970), Vol. IV, pp. 310-12.

[11] B. F. Skinner, "The Operational Analysis of Psychological Terms", *Psychological Review,* **52** (1945), 270-77.

[12] W. Sellars, "Empiricism and the Philosophy of Mind", *Minnesota Studies in the Philosophy of Science,* Vol. I, p. 328. Also in *Science, Perception, and Reality* (New York: The Humanities Press, 1963).
W. Sellars, "Thought and Action", in *Freedom & Determinism,* ed. by K. Lehrer (New York: Random House, 1966).
W. Sellars, "Intentionality and the Mental", *Minnesota Studies in the Philosophy of Science,* Vol. II (Appendix).

[13] Herbert A. Simon, "Motivational and Emotional Controls of Cognition", *Psychological Review,* **74** (1967), 29-39.