Chapter 23

Eliminative materialism and the propositional attitudes*

Paul M. Churchland

ELIMINATIVE materialism is the thesis that our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience. Our mutual understanding and even our introspection may then be reconstituted within the conceptual framework of completed neuroscience, a theory we may expect to be more powerful by far than the common-sense psychology it displaces, and more substantially integrated within physical science generally. My purpose in this paper is to explore these projections, especially as they bear on (1) the principal elements of common-sense psychology: the propositional attitudes (beliefs, desires, etc.), and (2) the conception of rationality in which these elements figure.

This focus represents a change in the fortunes of materialism. Twenty years ago, emotions, qualia, and 'raw feels' were held to be the principal stumbling blocks for the materialist program. With these barriers dissolving,[1] the locus of opposition has shifted. Now it is the realm of the intentional, the realm of the propositional attitude, that is most commonly held up as being both irreducible to and ineliminable in favor of anything from within a materialist framework. Whether and why this is so, we must examine.

Such an examination will make little sense, however, unless it is first appreciated that the relevant network of common-sense concepts does indeed constitute an empirical theory, with all the functions, virtues, *and perils* entailed by that status. I shall therefore begin with a brief sketch of this view and a summary rehearsal of its rationale. The resistance it encounters still surprises me. After all, common sense

Paul M. Churchland, 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy* 78 (1981).

* An earlier draft of this paper was presented at the University of Ottawa, and to the *Brain, Mind, and Person* colloquium at SUNY/Oswego. My thanks for the suggestions and criticisms that have informed the present version.

1. See Paul Feyerabend, 'Materialism and the Mind-Body Problem,' *Review* of *Metaphysics*, XVII. 1, 65 (September 1963): 49–66; Richard Rorty, 'Mind-Body Identity, Privacy, and Categories,' *ibid.*, XIX. 1, 73 (September 1965): 24–54; and my *Scientific Realism and the Plasticity of Mind* (New York: Cambridge, 1979).

has yielded up many theories. Recall the view that space has a preferred direction in which all things fall; that weight is an intrinsic feature of a body; that a force-free moving object will promptly return to rest; that the sphere of the heavens turns daily; and so on. These examples are clear, perhaps, but people seem willing to concede a theoretical component within common sense only if (1) the theory and the common sense involved are safely located in antiquity, and (2) the relevant theory is now so clearly false that its speculative nature is inescapable. Theories are indeed easier to discern under these circumstances. But the vision of hindsight is always 20/20. Let us aspire to some foresight for a change.

I. Why folk psychology is a theory

Seeing our common-sense conceptual framework for mental phenomena as a theory brings a simple and unifying organization to most of the major topics in the philosophy of mind, including the explanation and prediction of behavior, the semantics of mental predicates, action theory, the other-minds problem, the intentionality of mental states, the nature of introspection, and the mind-body problem. Any view that can pull this lot together deserves careful consideration.

Let us begin with the explanation of human (and animal) behavior. The fact is that the average person is able to explain, and even predict, the behavior of other persons with a facility and success that is remarkable. Such explanations and predictions standardly make reference to the desires, beliefs, fears, intentions, perceptions, and so forth, to which the agents are presumed subject. But explanations presuppose laws—rough and ready ones, at least—that connect the explanatory conditions with the behavior explained. The same is true for the making of predictions, and for the justification of subjunctive and counterfactual conditional concerning behavior. Reassuringly, a rich network of common-sense laws can indeed be reconstructed from this quotidean commerce of explanation and anticipation; its principles are familiar homilies; and their sundry functions are transparent. Each of us understands others, as well as we do, because we share a tacit command of an integrated body of lore concerning the law-like relations holding among external circumstances, internal states, and overt behavior. Given its nature and functions, this body of lore may quite aptly be called 'folk psychology.'[2]

This approach entails that the semantics of the terms in our familiar mentalistic vocabulary is to be understood in the same manner as the semantics of theoretical terms generally: the meaning of any theoretical term is fixed or constituted by the network of laws in which it figures. (This position is quite distinct from logical behaviorism. We deny that the relevant laws are analytic, and it is the lawlike

2. We shall examine a handful of these laws presently. For a more comprehensive sampling of the laws of folk psychology, see my *Scientific Realism and Plasticity of Mind, op. cit.*, ch. 4. For a detailed examination of the folk principles that underwrite action explanations in particular, see my 'The Logical Character of Action Explanations,' *Philosophical Review*, LXXIX, 2 (April 1970): 214–236.

connections generally that carry the semantic weight, not just the connections with overt behavior. But this view does account for what little plausibility logical behaviorism did enjoy.)

More importantly, the recognition that folk psychology is a theory provides a simple and decisive solution to an old skeptical problem, the problem of other minds. The problematic conviction that another individual is the subject of certain mental states is not inferred deductively from his behavior, nor is it inferred by inductive analogy from the perilously isolated instance of one's own case. Rather, that conviction is a singular *explanatory hypothesis* of a perfectly straightforward kind. Its function, in conjunction with the background laws of folk psychology, is to provide explanations/predictions/understanding of the individual's continuing behavior, and it is credible to the degree that it is successful in this regard over competing hypotheses. In the main, such hypotheses are successful, and so the belief that others enjoy the internal states comprehended by folk psychology is a reasonable belief.

Knowledge of other minds thus has no essential dependence on knowledge of one's own mind. Applying the principles of our folk psychology to our behavior, a Martian could justly ascribe to us the familiar run of mental states, even though his own psychology were very different from ours. He would not, therefore, be 'generalizing from his own case.'

As well, introspective judgments about one's own case turn out not to have any special status or integrity anyway. On the present view, an introspective judgment is just an instance of an acquired habit of conceptual response to one's internal states, and the integrity of any particular response is always contingent on the integrity of the acquired conceptual framework (theory) in which the response is framed. Accordingly, one's *introspective* certainty that one's mind is the seat of beliefs and desires may be as badly misplaced as was the classical man's *visual* certainty that the star-flecked sphere of the heavens turns daily.

Another conundrum is the intentionality of mental states. The 'propositional attitudes,' as Russell called them, form the systematic core of folk psychology; and their uniqueness and anomalous logical properties have inspired some to see here a fundamental contrast with anything that mere physical phenomena might conceivably display. The key to this matter lies again in the theoretical nature of folk psychology. The intentionality of mental states here emerges not as a mystery of nature, but as a structural feature of the concepts of folk psychology. Ironically, those same structural features reveal the very close affinity that folk psychology bears to theories in the physical sciences. Let me try to explain.

Consider the large variety of what might be called 'numerical attitudes' appearing in the conceptual framework of physical science: '. . . has a mass_{kg} of *n*', '. . . has a velocity of *n*', '. . . has a temperature_{K} of *n*', and so forth. These expressions are predicate-forming expressions: when one substitutes a singular term for a number into the place held by '*n*', a determinate predicate results. More interestingly, the relations between the various 'numerical attitudes' that result are precisely the

relations between the numbers 'contained' in those attitudes. More interesting still, the argument place that takes the singular terms for numbers is open to quantification. All this permits the expression of generalizations concerning the lawlike relations that hold between the various numerical attitudes in nature. Such laws involve quantification over numbers, and they exploit the mathematical relations holding in that domain. Thus, for example,

(1) $(x)(f)(m)[((x \text{ has a mass of } m) \,\&\, (x \text{ suffers a net force of } f)) \supset (x \text{ accelerates at } f/m)]$

Consider now the large variety of propositional attitudes: '. . . believes that *p*', '. . . desires that *p*', '. . . fears that *p*', '. . . is happy that *p*', etc. These expressions are predicate-forming expressions also. When one substitutes a singular term for a proposition into the place held by '*p*', a determinate predicate results, e.g., '. . . believes that Tom is tall.' (Sentences do not generally function as singular terms, but it is difficult to escape the idea that when a sentence occurs in the place held by '*p*', it is there functioning as or like a singular term. On this, more below.) More interestingly, the relations between the resulting propositional attitudes are characteristically the relations that hold between the propositions 'contained' in them, relations such as entailment, equivalence, and mutual inconsistency. More interesting still, the argument place that takes the singular terms for propositions is open to quantification. All this permits the expression of generalizations concerning the lawlike relations that hold among propositional attitudes. Such laws involve quantification over propositions, and they exploit various relations holding in that domain. Thus, for example,

(2) $(x)(p)[(x \text{ fears that } p) \supset (x \text{ desires that } \sim p)]$

(3) $(x)(p)[(x \text{ hopes that } p) \,\&\, (x \text{ discovers that } p)) \supset (x \text{ is pleased that } p)]$

(4) $(x)(p)(q)[((x \text{ believes that } p) \,\&\, (x \text{ believes that (if } p \text{ then } q))) \supset (\text{barring confusion, distraction, etc., } x \text{ believes that } q)]$

(5) $(x)(p)(q)[((x \text{ desires that } p) \,\&\, (x \text{ believes that (if } q \text{ then } p)) \,\&\, (x \text{ is able to bring it about that } q)) \supset (\text{barring conflicting desires or preferred strategies, } x \text{ brings it about that } q)]$[3]

3. Staying within an objectual interpretation of the quantifiers, perhaps the simplest way to make systematic sense of expressions like ⌜*x* believes that *p*⌝ and closed sentences formed therefrom is just to construe whatever occurs in the nested position held by '*p*', '*q*', etc. as there having the function of a singular term. Accordingly, the standard connectives, as they occur between terms in that nested position, must be construed as there functioning as operators that form compound singular terms from other singular terms, and not as sentence operators. The compound singular terms so formed denote the appropriate compound propositions. Substitutional quantification will of course underwrite a different interpretation, and there are other approaches as well. Especially appealing is the prosentential approach of Dorothy Grover, Joseph Camp, and Nuel Belnap, 'A Prosentential Theory of Truth,' *Philosophical Studies*, XXVII, 2 (February 1975): 73–125. But the resolution of these issues is not vital to the present discussion.

Not only is folk psychology a theory, it is so *obviously* a theory that it must be held a major mystery why it has taken until the last half of the twentieth century for philosophers to realize it. The structural features of folk psychology parallel perfectly those of mathematical physics; the only difference lies in the respective domain of abstract entities they exploit—numbers in the case of physics, and propositions in the case of psychology.

Finally, the realization that folk psychology is a theory puts a new light on the mind-body problem. The issue becomes a matter of how the ontology of one theory (folk psychology) is, or is not, going to be related to the ontology of another theory (completed neuroscience); and the major philosophical positions on the mind-body problem emerge as so many different anticipations of what future research will reveal about the intertheoretic status and integrity of folk psychology.

The identity theorist optimistically expects that folk psychology will be smoothly *reduced* by completed neuroscience, and its ontology preserved by dint of transtheoretic identities. The dualist expects that it will prove *ir*reducible to completed neuroscience, by dint of being a nonredundant description of an autonomous, nonphysical domain of natural phenomena. The functionalist also expects that it will prove irreducible, but on the quite different grounds that the internal economy characterized by folk psychology is not, in the last analysis, a law-governed economy of natural states, but an abstract organization of functional states, an organization instantiable in a variety of quite different material substrates. It is therefore irreducible to the principles peculiar to any of them.

Finally, the eliminative materialist is also pessimistic about the prospects for reduction, but his reason is that folk psychology is a radically inadequate account of our internal activities, too confused and too defective to win survival through intertheoretic reduction. On his view it will simply be displaced by a better theory of those activities.

Which of these fates is the real destiny of folk psychology, we shall attempt to divine presently. For now, the point to keep in mind is that we shall be exploring the fate of a theory, a systematic, corrigible, speculative *theory*.

II. Why folk psychology might (really) be false

Given that folk psychology is an empirical theory, it is at least an abstract possibility that its principles are radically false and that its ontology is an illusion. With the exception of eliminative materialism, however, none of the major positions takes this possibility seriously. None of them doubts the basic integrity or truth of folk psychology (hereafter, 'FP'), and all of them anticipate a future in which its laws and categories are conserved. This conservatism is not without some foundation. After all, FP does enjoy a substantial amount of explanatory and predictive success. And what better grounds than this for confidence in the integrity of its categories?

What better grounds indeed? Even so, the presumption in FP's favor is spurious,

born of innocence and tunnel vision. A more searching examination reveals a different picture. First, we must reckon not only with FP's successes, but with its explanatory failures, and with their extent and seriousness. Second, we must consider the long-term history of FP, its growth, fertility, and current promise of future development. And third, we must consider what sorts of theories are *likely* to be true of the etiology of our behavior, given what else we have learned about ourselves in recent history. That is, we must evaluate FP with regard to its coherence and continuity with fertile and well-established theories in adjacent and overlapping domains—with evolutionary theory, biology, and neuroscience, for example—because active coherence with the rest of what we presume to know is perhaps the final measure of any hypothesis.

A serious inventory of this sort reveals a very troubled situation, one which would evoke open skepticism in the case of any theory less familiar and dear to us. Let me sketch some relevant detail. When one centers one's attention not on what FP can explain, but on what it cannot explain or fails even to address, one discovers that there is a very great deal. As examples of central and important mental phenomena that remain largely or wholly mysterious within the framework of FP, consider the nature and dynamics of mental illness, the faculty of creative imagination, or the ground of intelligence differences between individuals. Consider our utter ignorance of the nature and psychological functions of sleep, that curious state in which a third of one's life is spent. Reflect on the common ability to catch an outfield fly ball on the run, or hit a moving car with a snowball. Consider the internal construction of a 3-D visual image from subtle differences in the 2-D array of stimulations in our respective retinas. Consider the rich variety of perceptual illusions, visual and otherwise. Or consider the miracle of memory, with its lightning capacity for relevant retrieval. On these and many other mental phenomena, FP sheds negligible light.

One particularly outstanding mystery is the nature of the learning process itself, especially where it involves large-scale conceptual change, and especially as it appears in its pre-linguistic or entirely nonlinguistic form (as in infants and animals), which is by far the most common form in nature. FP is faced with special difficulties here, since its conception of learning as the manipulation and storage of propositional attitudes founders on the fact that how to formulate, manipulate, and store a rich fabric of propositional attitudes is itself something that is learned, and is only one among many acquired cognitive skills. FP would thus appear constitutionally incapable of even addressing this most basic of mysteries.[4]

Failures on such a large scale do not (yet) show that FP is a false theory, but they

4. A possible response here is to insist that the cognitive activity of animals and infants is linguaformal in its elements, structures, and processing right from birth. J. A. Fodor, in *The Language of Thought* (New York: Crowell 1975), has erected a positive theory of thought on the assumption that the innate forms of cognitive activity have precisely the form here denied. For a critique of Fodor's view, see Patricia Churchland, 'Fodor on Language Learning,' *Synthese*, XXXVIII, 1 (May 1978): 149–159.

do move that prospect well into the range of real possibility, and they do show decisively that FP is *at best* a highly superficial theory, a partial and unpenetrating gloss on a deeper and more complex reality. Having reached this opinion, we may be forgiven for exploring the possibility that FP provides a positively misleading sketch of our internal kinematics and dynamics, one whose success is owed more to selective application and forced interpretation on our part than to genuine theoretical insight on FP's part.

A look at the history of FP does little to allay such fears, once raised. The story is one of retreat, infertility, and decadence. The presumed domain of FP used to be much larger than it is now. In primitive cultures, the behavior of most of the elements of nature were understood in intentional terms. The wind could know anger, the moon jealousy, the river generosity, the sea fury, and so forth. These were not metaphors. Sacrifices were made and auguries undertaken to placate or divine the changing passions of the gods. Despite its sterility, this animistic approach to nature has dominated our history, and it is only in the last two or three thousand years that we have restricted FP's literal application to the domain of the higher animals.

Even in this preferred domain, however, both the content and the success of FP have not advanced sensibly in two or three thousand years. The FP of the Greeks is essentially the FP we use today, and we are negligibly better at explaining human behavior in its terms than was Sophocles. This is a very long period of stagnation and infertility for any theory to display, especially when faced with such an enormous backlog of anomalies and mysteries in its own explanatory domain. Perfect theories, perhaps, have no need to evolve. But FP is profoundly imperfect. Its failure to develop its resources and extend its range of success is therefore darkly curious, and one must query the integrity of its basic categories. To use Imre Lakatos' terms, FP is a stagnant or degenerating research program, and has been for millennia.

Explanatory success to date is of course not the only dimension in which a theory can display virtue or promise. A troubled or stagnant theory may merit patience and solicitude on other grounds; for example, on grounds that it is the only theory or theoretical approach that fits well with other theories about adjacent subject matters, or the only one that promises to reduce to or be explained by some established background theory whose domain encompasses the domain of the theory at issue. In sum, it may rate credence because it holds promise of theoretical integration. How does FP rate in this dimension?

It is just here, perhaps, that FP fares poorest of all. If we approach *homo sapiens* from the perspective of natural history and the physical sciences, we can tell a coherent story of his constitution, development, and behavioral capacities which encompasses particle physics, atomic and molecular theory, organic chemistry, evolutionary theory, biology, physiology, and materialistic neuroscience. That story, though still radically incomplete, is already extremely powerful, outperforming FP at many points even in its own domain. And it is deliberately and self-consciously coherent with the rest of our developing world picture. In short, the greatest theor-

etical synthesis in the history of the human race is currently in our hands, and parts of it already provide searching descriptions and explanations of human sensory input, neural activity, and motor control.

But FP is no part of this growing synthesis. Its intentional categories stand magnificently alone, without visible prospect of reduction to that larger corpus. A successful reduction cannot be ruled out, in my view, but FP's explanatory impotence and long stagnation inspire little faith that its categories will find themselves neatly reflected in the framework of neuroscience. On the contrary, one is reminded of how alchemy must have looked as elemental chemistry was taking form, how Aristotelean cosmology must have looked as classical mechanics was being articulated, or how the vitalist conception of life must have looked as organic chemistry marched forward.

In sketching a fair summary of this situation, we must make a special effort to abstract from the fact that FP is a central part of our current *lebenswelt*, and serves as the principal vehicle of our interpersonal commerce. For these facts provide FP with a conceptual inertia that goes far beyond its purely theoretical virtues. Restricting ourselves to this latter dimension, what we must say is that FP suffers explanatory failures on an epic scale, that it has been stagnant for at least twenty-five centuries, and that its categories appear (so far) to be incommensurable with or orthogonal to the categories of the background physical science whose long-term claim to explain human behavior seems undeniable. Any theory that meets this description must be allowed a serious candidate for outright elimination.

We can of course insist on no stronger conclusion at this stage. Nor is it my concern to do so. We are here exploring a possibility, and the facts demand no more, and no less, than it be taken seriously. The distinguishing feature of the eliminative materialist is that he takes it very seriously indeed.

III. Arguments against elimination

Thus the basic rationale of eliminative materialism: FP is a theory, and quite probably a false one; let us attempt, therefore to transcend it.

The rationale is clear and simple, but many find it uncompelling. It will be objected that FP is not, strictly speaking, an *empirical* theory; that it is not false, or at least not refutable by empirical considerations; and that it ought not or cannot be transcended in the fashion of a defunct empirical theory. In what follows we shall examine these objections as they flow from the most popular and best-founded of the competing positions in the philosophy of mind: functionalism.

An antipathy toward eliminative materialism arises from two distinct threads running through contemporary functionalism. The first thread concerns the *normative* character of FP, or at least of that central core of FP which treats of the propositional attitudes. FP, some will say, is a characterization of an ideal, or at least praiseworthy mode of internal activity. It outlines not only what it is to have and

process beliefs and desires, but also (and inevitably) what it is to be rational in their administration. The ideal laid down by FP may be imperfectly achieved by empirical humans, but this does not impugn FP as a normative characterization. Nor need such failures seriously impugn FP even as a descriptive characterization, for it remains true that our activities can be both usefully and accurately understood as rational *except for* the occasional lapse due to noise, interference, or other breakdown, which defects empirical research may eventually unravel. Accordingly, though neuroscience may usefully augment it, FP has no pressing need to be displaced, even as a descriptive theory; nor could it be replaced, qua normative characterization, by any descriptive theory of neural mechanisms, since rationality is defined over propositional attitudes like beliefs and desires. FP, therefore, is here to stay.

Daniel Dennett has defended a view along these lines.[5] And the view just outlined gives voice to a theme of the property dualists as well. Karl Popper and Joseph Margolis both cite the normative nature of mental and linguistic activity as a bar to their penetration or elimination by any descriptive/materialist theory.[6] I hope to deflate the appeal of such moves below.

The second thread concerns the *abstract* nature of FP. The central claim of functionalism is that the principles of FP characterize our internal states in a fashion that makes no reference to their intrinsic nature or physical constitution. Rather, they are characterized in terms of the network of causal relations they bear to one another, and to sensory circumstances and overt behavior. Given its abstract specification, that internal economy may therefore be realized in a nomically heterogeneous variety of physical systems. All of them may differ, even radically, in their physical constitution, and yet at another level, they will all share the same nature. This view, says Fodor, 'is compatible with very strong claims about the ineliminabilty of mental language from behavioral theories.'[7] Given the real possibility of multiple instantiations in heterogeneous physical substrates, we cannot eliminate the functional characterization in favor of any theory peculiar to one such substrate. That would preclude our being able to describe the (abstract) organization that any one instantiation shares with all the other. A functional characterization of our internal states is therefore here to stay.

This second theme, like the first, assigns a faintly stipulative character to FP, as if the onus were on the empirical systems to instantiate faithfully the organization that FP specifies, instead of the onus being on FP to describe faithfully the internal activities of a naturally distinct class of empirical systems. This impression is enhanced by the standard examples used to illustrate the claims of functionalism—

5. Most explicitly in 'Three Kinds of Intentional Psychology', in R. Healy, ed. *Reduction, Time, and Reality* (Cambridge: Cambridge University Press, 1975): 37–60; (See Chapter 19 of this volume), but this theme of Dennett's goes all the way back to his 'Intentional Systems,' this JOURNAL, LXVIII, 4 (Feb. 25, 1971): 87–106; reprinted in his *Brainstorms* (Montgomery, Vt.: Bradford Books, 1978).
6. Popper, *Objective Knowledge* (New York: Oxford, 1972); with J. Eccles, *The Self and Its Brain* (New York: Springer Verlag, 1978). Margolis, *Persons and Minds* (Boston: Reidel, 1978).
7. *Psychological Explanation* (New York: Random House, 1968), p. 116.

mousetraps, valve-lifters, arithmetical calculators, computers, robots, and the like. These are artifacts, constructed to fill a preconceived bill. In such cases, a failure of fit between the physical system and the relevant functional characterization impugns only the former, not the latter. The functional characterization is thus removed from empirical criticism in a way that is most unlike the case of an empirical theory. One prominent functionalist—Hilary Putnam—has argued outright that FP is not a corrigible theory at all.[8] Plainly, if FP is construed on these models, as regularly it is, the question of its empirical integrity is unlikely ever to pose itself, let alone receive a critical answer.

Although fair to some functionalists, the preceding is not entirely fair to Fodor. On his view the aim of psychology is to find the *best* functional characterization of ourselves, and what that is remains an empirical question. As well, his argument for the ineliminability of mental vocabulary from psychology does not pick out current FP in particular as ineliminable. It need claim only that *some* abstract functional characterization must be retained, some articulation or refinement of FP perhaps.

His estimate of eliminative materialism remains low, however. First, it is plain that Fodor thinks there is nothing fundamentally or interestingly wrong with FP. On the contrary, FP's central conception of cognitive activity—as consisting in the manipulation of propositional attitudes—turns up as the central element in Fodor's own theory on the nature of thought (*The Language of Thought, op. cit.*). And second, there remains the point that, whatever tidying up FP may or may not require, it cannot be displaced by any naturalistic theory of our physical substrate, since it is the abstract functional features of his internal states that make a person, not the chemistry of his substrate.

All of this is appealing. But almost none of it, I think, is right. Functionalism has too long enjoyed its reputation as a daring and *avant garde* position. It needs to be revealed for the short-sighted and reactionary position it is.

IV. The conservative nature of functionalism

A valuable perspective on functionalism can be gained from the following story. To begin with, recall the alchemists' theory of inanimate matter. We have here a long and variegated tradition, of course, not a single theory, but our purposes will be served by a gloss.

The alchemists conceived the 'inanimate' as entirely continuous with animated matter, in that the sensible and behavioral properties of the various substances are owed to the ensoulment of baser matter by various spirits or essences. These non-material aspects were held to undergo development, just as we find growth and development in the various souls of plants, animals, and humans. The alchemist's

8. 'Robots: Machines or Artificially Created Life?', this JOURNAL, LXI, 21 (Nov. 12, 1964): 668–691, pp. 675, 681 ff.

peculiar skill lay in knowing how to seed, nourish, and bring to maturity the desired spirits enmattered in the appropriate combinations.

On one orthodoxy, the four fundamental spirits (for 'inanimate' matter) were named 'mercury,' 'sulphur,' 'yellow arsenic,' and 'sal ammoniac.' Each of these spirits was held responsible for a rough but characteristic syndrome of sensible, combinatorial, and causal properties. The spirit mercury, for example, was held responsible for certain features typical of metallic substances—their shininess, liquefiability, and so forth. Sulphur was held responsible for certain residual features typical of metals, and for those displayed by the ores from which running metal could be distilled. Any given metallic substance was a critical orchestration principally of these two spirits. A similar story held for the other two spirits, and among the four of them a certain domain of physical features and transformations was rendered intelligible and controllable.

The degree of control was always limited, of course. Or better, such prediction and control as the alchemists possessed was owed more to the manipulative lore acquired as an apprentice to a master, than to any genuine insight supplied by the theory. The theory followed, more than it dictated, practice. But the theory did supply some rhyme to the practice, and in the absence of a developed alternative it was sufficiently compelling to sustain a long and stubborn tradition.

The tradition had become faded and fragmented by the time the elemental chemistry of Lavoisier and Dalton arose to replace it for good. But let us suppose that it had hung on a little longer—perhaps because the four-spirit orthodoxy had become a thumbworn part of everyman's common sense—and let us examine the nature of the conflict between the two theories and some possible avenues of resolution.

No doubt the simplest line of resolution, and the one which historically took place, is outright displacement. The dualistic interpretation of the four essences—as immaterial spirits—will appear both feckless and unnecessary given the power of the corpuscularian taxonomy of atomic chemistry. And a reduction of the old taxonomy to the new will appear impossible, given the extent to which the comparatively toothless old theory cross-classifies things relative to the new. Elimination would thus appear the only alternative—*unless* some cunning and determined defender of the alchemical vision has the wit to suggest the following defense.

Being 'ensouled by mercury,' or 'sulphur,' or either of the other two so-called spirits, is actually a *functional* state. The first, for example, is defined by the disposition to reflect light, to liquefy under heat, to unite with other matter in the same state, and so forth. And each of these four states is related to the others, in that the syndrome for each varies as a function of which of the other three states is also instantiated in the same substrate. Thus the level of description comprehended by the alchemical vocabulary is abstract: various material substances, suitably 'ensouled,' can display the features of a metal, for example, or even of gold specifically. For it is the total syndrome of occurrent and causal properties which matters,

not the corpuscularian details of the substrate. Alchemy, it is concluded, comprehends a level of organization in reality distinct from and irreducible to the organization found at the level of corpuscularian chemistry.

This view might have had considerable appeal. After all, it spares alchemists the burden of defending immaterial souls that come and go; it frees them from having to meet the very strong demands of a naturalistic reduction; and it spares them the shock and confusion of outright elimination. Alchemical theory emerges as basically all right! Nor need they appear too obviously stubborn or dogmatic in this. Alchemy as it stands, they concede, may need substantial tidying up, and experience must be our guide. But we need not fear its naturalistic displacement, they remind us, since it is the particular orchestration of the syndromes of occurrent and causal properties which makes a piece of matter gold, not the idiosyncratic details of its corpuscularian substrate. A further circumstance would have made this claim even more plausible. For the fact is, the alchemists *did* know how to make gold, in this relevantly weakened sense of 'gold', and they could do so in a variety of ways. Their 'gold' was never as perfect, alas, as the 'gold' nurtured in nature's womb, but what mortal can expect to match the skills of nature herself?

What this story shows is that it is at least possible for the constellation of moves, claims, and defenses characteristic of functionalism to constitute an outrage against reason and truth, and to do so with a plausibility that is frightening. Alchemy is a terrible theory, well-deserving of its complete elimination, and the defense of it just explored is reactionary, obfuscatory, retrograde, and wrong. But in historical context, that defense might have seemed wholly sensible, even to reasonable people.

The alchemical example is a deliberately transparent case of what might well be called 'the functionalist strategem,' and other cases are easy to imagine. A cracking good defense of the phlogiston theory of combustion can also be constructed along these lines. Construe being highly phlogisticated and being dephlogisticated as functional states defined by certain syndromes of causal dispositions; point to the great variety of natural substrates capable of combustion and calxification; claim an irreducible functional integrity for what has proved to lack any natural integrity; and bury the remaining defects under a pledge to contrive improvements. A similar recipe will provide new life for the four humors of medieval medicine, for the vital essence or archeus of pre-modern biology, and so forth.

If its application in these other cases is any guide, the functionalist strategem is a smokescreen for the preservation of error and confusion. Whence derives our assurance that in contemporary journals the same charade is not being played out on behalf of FP? The parallel with the case of alchemy is in all other respects distressingly complete, right down to the parallel between the search for artificial gold and the search for artificial intelligence!

Let me not be misunderstood on this last point. Both aims are worthy aims: thanks to nuclear physics, artificial (but real) gold is finally within our means, if only in submicroscopic quantities; and artificial (but real) intelligence eventually will be. But just as the careful orchestration of superficial syndromes was the wrong

way to produce genuine gold, so may the careful orchestration of superficial syndromes be the wrong way to produce genuine intelligence. Just as with gold, what may be required is that our science penetrate to the underlying *natural* kind that gives rise to the total syndrome directly.

In summary, when confronted with the explanatory impotence, stagnant history, and systematic isolation of the intentional idioms of FP, it is not an adequate or responsive defense to insist that those idioms are abstract, functional, and irreducible in character. For one thing, this same defense could have been mounted with comparable plausibility no matter *what* haywire network of internal states our folklore had ascribed to us. And for another, the defense assumes essentially what is at issue: it assumes that it is the intentional idioms of FP, plus or minus a bit, that express the *important* features shared by all cognitive systems. But they may not. Certainly it is wrong to assume that they do, and then argue against the possibility of a materialistic displacement on grounds that it must descibe matters at a level that is different from the important level. This just begs the question in favor of the older framework.

Finally, it is very important to point out that eliminative materialism is strictly *consistent* with the claim that the essence of a cognitive system resides in the abstract functional organization of its internal states. The eliminative materialist is not committed to the idea that the correct account of cognition *must* be a naturalistic account, though he may be forgiven for exploring the possibility. What he does hold is that the correct account of cognition, whether functionalistic or naturalistic, will bear about as much resemblance to FP as modern chemistry bears to four-spirit alchemy.

Let us now try to deal with the argument, against eliminative materialism, from the normative dimension of FP. This can be dealt with rather swiftly, I believe.

First, the fact that the regularities ascribed by the intentional core of FP are predicated on certain logical relations among propositions is not by itself grounds for claiming anything essentially normative about FP. To draw a relevant parallel, the fact that the regularities ascribed by the classical gas law are predicated on arithmetical relations between numbers does not imply anything essentially normative about the classical gas law. And logical relations between propositions are as much an objective matter of abstract fact as are arithmetical relations between numbers. In this respect, the law

(4) $(x)(p)(q)$ [((x believes that p) & (x believes that (if p then q)))
$\supset$ (barring confusion, distraction, etc., x believes that q)]

is entirely on a par with the classical gas law

(6) $(x)(P)(V)(\mu)$[((x has a pressure P) & (x has a volume V)
& (x has a quantity μ)) $\supset$ (barring very high pressure or density,
x has a temperature of $PV/\mu R$)]

A normative dimension enters only because we happen to *value* most of the patterns ascribed by FP. But we do not value all of them. Consider

(7) $(x)(p)[((x \text{ desires with all his heart that } p) \,\&\, (x \text{ learns that } \sim p)) \supset (\text{barring unusual strength of character. } x \text{ is shattered that } \sim p)]$

Moreover, and as with normative convictions generally, fresh insight may motivate major changes in what we value.

Second, the laws of FP ascribe to us only a very minimal and truncated rationality, not an ideal rationality as some have suggested. The rationality characterized by the set of all FP laws falls well short of an ideal rationality. This is not surprising. We have no clear or finished conception of ideal rationality anyway; certainly the ordinary man does not. Accordingly, it is just not plausible to suppose that the explanatory failures from which FP suffers are owed primarily to human failure to live up to the ideal standard it provides. Quite to the contrary, the conception of rationality it provides appears limping and superficial, especially when compared with the dialectical complexity of our scientific history, or with the ratiocinative virtuosity displayed by any child.

Third, even if our current conception of rationality—and more generally, of cognitive virtue—is largely constituted within the sentential/propositional framework of FP, there is no guarantee that this framework is adequate to the deeper and more accurate account of cognitive virtue which is clearly needed. Even if we concede the categorial integrity of FP, at least as applied to language-using humans, it remains far from clear that the basic parameters of intellectual virtue are to be found at the categorial level comprehended by the propositional attitudes. After all, language use is something that is learned, by a brain already capable of vigorous cognitive activity; language use is acquired as only one among a great variety of learned manipulative skills; and it is mastered by a brain that evolution has shaped for a great many functions, language use being only the very latest and perhaps the least of them. Against the background of these facts, language use appears as an extremely peripheral activity, as a racially idiosyncratic mode of social interaction which is mastered thanks to the versatility and power of a more basic mode of activity. Why accept then, a theory of cognitive activity that models its elements on the elements of human language? And why assume that the fundamental parameters of intellectual virtue are or can be defined over the elements at this superficial level?

A serious advance in our appreciation of cognitive virtue would thus seem to *require* that we go beyond FP, that we transcend the poverty of FP's conception of rationality by transcending its propositional kinematics entirely, by developing a deeper and more general kinematics of cognitive activity, and by distinguishing within this new framework which of the kinematically possible modes of activity are to be valued and encouraged (as more efficient, reliable, productive, or whatever). Eliminative materialism thus does not imply the end of our normative concerns. It implies only that they will have to be reconstituted at a more revealing level of understanding, the level that a matured neuroscience will provide.

What a theoretically informed future might hold in store for us, we shall now turn to explore. Not because we can foresee matters with any special clarity, but because it is important to try to break the grip on our imagination held by the propositional kinematics of FP. As far as the present section is concerned, we may summarize our conclusions as follows. FP is nothing more and nothing less than a culturally entrenched theory of how we and the higher animals work. It has no special features that make it empirically invulnerable, no unique functions that make it irreplaceable, no special status of any kind whatsoever. We shall turn a skeptical ear then, to any special pleading on its behalf.

V. Beyond folk psychology

What might the elimination of FP actually involve—not just the comparatively straightforward idioms for sensation, but the entire apparatus of propositional attitudes? That depends heavily on what neuroscience might discover, and on our determination to capitalize on it. Here follow three scenarios in which the operative conception of cognitive activity is progressively divorced from the forms and categories that characterize natural language. If the reader will indulge the lack of actual substance, I shall try to sketch some plausible form.

First suppose that research into the structure and activity of the brain, both fine-grained and global, finally does yield a new kinematics and correlative dynamics for what is now thought of as cognitive activity. The theory is uniform for all terrestrial brains, not just human brains, and it makes suitable conceptual contact with both evolutionary biology and non-equilibrium thermodynamics. It ascribes to us, at any given time, a set or configuration of complex states, which are specified within the theory as figurative 'solids' within a four- or five-dimensional phase space. The laws of the theory govern the interaction, motion, and transformation of these 'solid' states within that space, and also their relations to whatever sensory and motor transducers the system possesses. As with celestial mechanics, the exact specification of the 'solids' involved and the exhaustive accounting of all dynamically relevant adjacent 'solids' is not practically possible, for many reasons, but here also it turns out that the obvious approximations we fall back on yield excellent explanations/predictions of internal change and external behavior; at least in the short term. Regarding long-term activity, the theory provides powerful and unified accounts of the learning process, the nature of mental illness, and variations in character and intelligence across the animal kingdom as well as across individual humans.

Moreover, it provides a straightforward account of 'knowledge,' as traditionally conceived. According to the new theory, any declarative sentence to which a speaker would give confident assent is merely a one-dimensional *projection*—through the compound lens of Wernicke's and Broca's areas onto the idiosyncratic surface of the speaker's language—a one-dimensional projection of a four- or five-

dimensional 'solid' that is an element in his true kinematical state. (Recall the shadows on the wall of Plato's cave.) Being projections of that inner reality, such sentences do carry significant information regarding it and are thus fit to function as elements in a communication system. On the other hand, being *sub*dimensional projections, they reflect but a narrow part of the reality projected. They are therefore *un*fit to represent the deeper reality in all its kinematically, dynamically, and even normatively relevant respects. That is to say, a system of propositional attitudes, such as FP, must inevitably fail to capture what is going on here, though it may reflect just enough superficial structure to sustain an alchemylike tradition among folk who lack any better theory. From the perspective of the newer theory, however, it is plain that there simply are no law-governed states of the kind FP postulates. The real laws governing our internal activities are defined over different and much more complex kinematical states and configurations, as are the normative criteria for developmental integrity and intellectual virtue.

A theoretical outcome of the kind just described may fairly be counted as a case of elimination of one theoretical ontology in favor of another, but the success here imagined for systematic neuroscience need not have any sensible effect on common practice. Old ways die hard, and in the absence of some practical necessity, they may not die at all. Even so, it is not inconceivable that some segment of the population, or all of it, should become intimately familiar with the vocabulary required to characterize our kinematical states, learn the laws governing their interactions and behavioral projections, acquire a facility in their first-person ascription, and displace the use of FP altogether, even in the marketplace. The demise of FP's ontology would then be complete.

We may now explore a second and rather more radical possibility. Everyone is familiar with Chomsky's thesis that the human mind or brain contains innately and uniquely the abstract structures for learning and using specifically human natural languages. A competing hypothesis is that our brain does indeed contain innate structures, but that those structures have as their original and still primary function the organization of perceptual experience, the administration of linguistic categories being an acquired and additional function for which evolution has only incidentally suited them.[9] This hypothesis has the advantage of not requiring the evolutionary saltation that Chomsky's view would seem to require, and there are other advantages as well. But these matters need not concern us here. Suppose, for our purposes, that this competing view is true, and consider the following story.

Research into the neural structures that fund the organization and processing of perceptual information reveals that they are capable of administering a great variety of complex tasks, some of them showing a complexity far in excess of that shown by natural language. Natural languages, it turns out, exploit only a very elementary portion of the available machinery, the bulk of which serves far more

9. Richard Gregory defends such a view in 'The Grammar of Vision,' *Listener*, LXXXIII, 2133 (February 1970): 242–246; reprinted in his *Concepts and Mechanisms of Perception* (London: Duckworth, 1975), pp. 622–629.

complex activities beyond the ken of the propositional conceptions of FP. The detailed unraveling of what that machinery is and of the capacities it has makes it plain that a form of language far more sophisticated than 'natural' language, though decidedly 'alien' in its syntactic and semantic structures, could also be learned and used by our innate systems. Such a novel system of communication, it is quickly realized, could raise the efficiency of information exchange between brains by an order of magnitude, and would enhance epistemic evaluation by a comparable amount, since it would reflect the underlying structure of our cognitive activities in greater detail than does natural language.

Guided by our new understanding of those internal structures, we manage to construct a new system of verbal communication entirely distinct from natural language, with a new and more powerful combinatorial grammar over novel elements forming novel combinations with exotic properties. The compounded strings of this alternative system—call them 'übersatzen'—are not evaluated as true or false, nor are the relations between them remotely analogous to the relations of entailment, etc., that hold between sentences. They display a different organization and manifest different virtues.

Once constructed, this 'language' proves to be learnable; it has the power projected; and in two generations it has swept the planet. Everyone uses the new system. The syntactic forms and semantic categories of so-called 'natural' language disappear entirely. And with them disappear the propositional attitudes of FP, displaced by a more revealing scheme in which (of course) 'übersatzenal attitudes' play the leading role. FP again suffers elimination.

This second story, note, illustrates a theme with endless variations. There are possible as many different 'folk psychologies' as there are possible differently structured communication systems to serve as models for them.

A third and even stranger possibility can be outlined as follows. We know that there is considerable lateralization of function between the two cerebral hemispheres, and that the two hemispheres make use of the information they get from each other by way of the great cerebral commissure—the corpus callosum—a giant cable of neurons connecting them. Patients whose commissure has been surgically severed display a variety of behavioral deficits that indicate a loss of access by one hemisphere to information it used to get from the other. However, in people with callosal agenesis (a congenital defect in which the connecting cable is simply absent), there is little or no behavioral deficit, suggesting that the two hemisphere have learned to exploit the information carried in other less direct pathways connecting them through the subcortical regions. This suggests that, even in the normal case, a developing hemisphere *learns* to make use of the information the cerebral commissure deposits at its doorstep. What we have then, in the case of a normal human, is two physically distinct cognitive systems (both capable of independent function) responding in a systematic and learned fashion to exchanged information. And what is especially interesting about this case is the sheer amount of information exchanged. The cable of the commissure consists of

≈200 million neurons,[10] and even if we assume that each of these fibres is capable of one of only two possible states each second (a most conservative estimate), we are looking at a channel whose information capacity is $> 2 \times 10^8$ binary bits/second. Compare this to the < 500 bits/second capacity of spoken English.

Now, if two distinct hemispheres can learn to communicate on so impressive a scale, why shouldn't two distinct brains learn to do it also? This would require an artificial 'commissure' of some kind, but let us suppose that we can fashion a workable transducer for implantation at some site in the brain that research reveals to be suitable, a transducer to convert a symphony of neural activity into (say) microwaves radiated from an aerial in the forehead, and to perform the reverse function of converting received microwaves back into neural activation. Connecting it up need not be an insuperable problem. We simply trick the normal processes of dendretic arborization into growing their own myriad connections with the active microsurface of the transducer.

Once the channel is opened between two or more people, they can learn (*learn*) to exchange information and coordinate their behavior with the same intimacy and virtuosity displayed by your own cerebral hemispheres. Think what this might do for hockey teams, and ballet companies, and research teams! If the entire population were thus fitted out, spoken language of any kind might well disappear completely, a victim of the 'why crawl when you can fly?' principle. Libraries become filled not with books, but with long recordings of exemplary bouts of neural activity. These constitute a growing cultural heritage, an evolving 'Third World,' to use Karl Popper's terms. But they do not consist of sentences or arguments.

How will such people understand and conceive of other individuals? To this question I can only answer, 'In roughly the same fashion that your right hemisphere 'understands' and 'conceives of' your left hemisphere—intimately and efficiently, but not propositionally!'

These speculations, I hope, will evoke the required sense of untapped possibilities, and I shall in any case bring them to a close here. Their function is to make some inroads into the aura of inconceivability that commonly surrounds the idea that we might reject FP. The felt conceptual strain even finds expression in an argument to the effect that the thesis of eliminative materialism is incoherent since it denies the very conditions presupposed by the assumption that it is meaningful. I shall close with a brief discussion of this very popular move.

As I have received it, the reductio proceeds by pointing out that the statement of eliminative materialism is just a meaningless string of marks or noises, unless that string is the expression of a certain *belief*, and a certain *intention* to communicate, and a *knowledge* of the grammar of the language, and so forth. But if the statement of eliminative materialism is true, then there are no such states to express. The statement at issue would then be a meaningless string of marks or noises. It would therefore *not* be true. Therefore it is not true. Q.E.D.

10. M. S. Gazzaniga and J. E. LeDoux, *The Integrated Mind* (New York: Plenum Press, 1975).

The difficulty with any nonformal reductio is that the conclusion against the initial assumption is always no better than the material assumptions invoked to reach the incoherent conclusion. In this case the additional assumptions involve a certain theory of meaning, one that presupposes the integrity of FP. But formally speaking, one can as well infer, from the incoherent result, that this theory of meaning is what must be rejected. Given the independent critique of FP leveled earlier, this would even seem the preferred option. But in any case, one cannot simply assume that particular theory of meaning without begging the question at issue, namely, the integrity of FP.

The question-begging nature of this move is most graphically illustrated by the following analogue, which I owe to Patricia Churchland.[11] The issue here, placed in the seventeenth century, is whether there exists such a substance as *vital spirit.* At the time, this substance was held, without significant awareness of real alternatives, to be that which distinguished the animate from the inanimate. Given the monopoly enjoyed by this conception, given the degree to which it was integrated with many of our other conceptions, and given the magnitude of the revisions any serious alternative conception would require, the following refutation of any anti-vitalist claim would be found instantly plausible.

> The anti-vitalist says that there is no such thing as vital spirit. But this claim is self-refuting. The speaker can expect to be taken seriously only if his claim cannot. For if the claim is true, then the speaker does not have vital spirit and must be *dead.* But if he is dead, then his statement is a meaningless string of noises. devoid of reason and truth.

The question-begging nature of this argument does not. I assume, require elaboration. To those moved by the earlier argument. I commend the parallel for examination.

The thesis of this paper may be summarized as follows. The propositional attitudes of folk psychology do not constitute an unbreachable barrier to the advancing tide of neuroscience. On the contrary, the principled displacement of folk psychology is not only richly possible, it represents one of the most intriguing theoretical displacements we can currently imagine.

11. 'Is Determinism Self-Refuting?', *Mind* 90 (1981): 99–101.