INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9207096

Clinical implications of the mind/body question: A critique of mechanism

Lipson, Michael Aaron, Ph.D.

City University of New York, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A

CLINICAL IMPLICATIONS OF THE MIND/BODY QUESTION: A CRITIQUE OF MECHANISM

by

Michael A. Lipson

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1991

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

June 27, 1991
Date

Chair of Examining Committee

July 1, 1991
Date

Executive Officer

I. H. Paul, PhD

Vera Paster, PhD

Michael Moskowitz, PhD
Supervisory Committee

The City University of New York

Clinical Implications of the Mind/Body Question: A Critique of Mechanism

Table of Contents

Introduction

Much of current psychology and psychiatry suffers from a discrepancy between the claims of prevalent theories and the facts of clinical psychotherapeutic practice. The dominant view of the human mind on entering the 1990's has continued the twentieth century reduction of mental events and contents, regarding them as products of physical states and mechanistic causes (Robinson, 1986). Psychotherapy, meanwhile, rests in varying degrees on the encouragement of the patient's power of choice, production of meaning, capacity for insight, and investment in human relationship (Yalom, 1980; Mitchell, 1988; Schorr, 1984). All these therapeutically encouraged capacities would be of merely derivative or secondary status, if not outright illusions, in a mechanistic psychological system. The source of this discrepancy lies deep in the materialist bent of Western medicine and psychology.

Recent developments in the neurosciences (Kolb, 1985; McClelland et al., 1986) and computer science (Gardner, 1985; Minsky, 1985; Johnson-Laird, 1988) have made the ancient mind/body problem pressing once again at a

theoretical level, while the new field of psychoneuroimmunology (Pelletier, 1977; Ader, 1981) brings new clinical relevance to the question. In fact, not only these new trends, but every clinical encounter takes at least an implicit stance toward the mind/body problem. Psychotherapy therefore could profit from the realignment of its often superb clinical practice with more apposite theoretical foundations. The aim of this dissertation is to contribute to such a realignment by demonstrating that some functions of consciousness (for example, the understanding) are fundamental elements of psychological science, and cannot be reduced to or derived from other (for example, neurological) functions.

Overview

Beginning with selected moments in the history of mind/body problematization, the first chapter will examine Aristotle's tripartite view of man. Aristotle in De Anima offered a picture of the human constitution that has relevance as a possible alternative to the materialist monism of our own time. Specifically, he suggested that within the psyche there must be a radically un-determined, formless function to account for the act of self-observation that is inherent in psychology and psychotherapy. Moments

in the subsequent history of mind/body theory are also traced, including Descartes' formulation, its significance, and the development of eliminative materialism in this century. This investigation suggests that at each stage in the post-Renaissance history of psychology, a physicalist bargain has been struck, such that when some new area of psychic functioning is allowed to receive scientific attention, it is always done under cover of a mechanistic theory that likens psychic functioning to, or derives it from, the physically observable world of natural scientific observation. Instances of the physicalist bargain in the recent history of psychology highlight this phenomenon.

The second chapter looks more closely at the logical problem of certain kinds of self-reference or reflexivity. Drawing on the work of Hans Jonas (1987), Georg Kuehlewind (1982) and Lynne Ruder Baker (1987), I show that any strong mechanistic psychology contradicts itself. This conclusion, the heart of the thesis, implies the necessary ascription of primary reality and causation to at least <u>some</u> aspects of mental functioning. Arguments for and against physicalism and mechanism in psychology are closely analyzed. The mathematical/logical form of this question, as presented in the work of K. Goedel, is also discussed in its relevance to the mind/body debate.

The third chapter looks at the therapeutic implications of the central thesis of psychological irreducibility. They include a renewed understanding of such concepts as rapport and autonomy, as well as suggestions for the extension of psychotherapy as a development of psychological capacity rather than performance. It is capacities, rather than finished mental products, that must be considered the primary realities of psychological science. This perspective suggests a change in the understanding of data from psychoneuroimmunological studies, new avenues for mind/body healing, and a theoretical foundation for work with the terminally ill. Finally, the anti-reductive thesis recognizes as real the responsibility of the patient as subject.

Implications of Mechanism

It is typical for current college textbooks of psychology to begin with chapters on neuroanatomy, and to include statements such as the following:

> Your feelings and thoughts, your beliefs and dreams, your plans and actions are all the product of activity in your central nervous system, of which the brain is the main component. (Wortmann et al., 1988)

As the recent, highly successful public television series The Mind put it, "the mind is what the brain does." This view determines, in the medical and psychological

communities, what research receives funding, how drugs are developed and prescribed, how therapists and patients view themselves, and even to some degree how psychotherapeutic treatment proceeds. Profoundly misguided approaches to alcohol and drug abuse (c.f Peele, 1985) are among the many consequences of mechanistic bias.

For purposes of this thesis, I define as mechanistic all theories of psychological functioning that locate the causes of conscious mental events exclusively in <u>other</u> realms -- that is, in anything non-conscious, non-mental, or non-eventful. Examples of these other realms include neurological events, unconscious drive states, stimulus-response patterns, and computerlike "programs." In general, individual mechanistic systems go much further than mere reduction, sometimes eliminating conscious mental events altogether, as in theoretical behaviorism. In most mechanisms (<u>e.g.</u> in physicalism), the further attempt is made to derive conscious mental events from a finite number of natural scientific laws or rules. They therefore tend to follow deterministic or probabilistic principles. Whether a psychological theory subscribes to physical law (as in the case of eliminative materialism) or remains non-physical (as in the case of some cognitive scientists), it is "mechanistic" unless it allows for mental capacities beyond

formal delineation. For example, to the extent that psychoanalysis adheres to the principle of psychic determinism, it is also a mechanistic system.

It is worth emphasizing that the definition of mechanism given above is intentionally broad, because the critique exercised in the present thesis applies not only to existing, but also to all possible mechanisms. The procedure at hand amounts, therefore, to something like a contrast between styles, rather than between specific scientific programs. Mechanisms are various, but have one common epistemological style; perhaps there is a similar variety-in-unity on the other side of the dichotomy. This other style, of non-mechanistic psychologies, is difficult not only to describe, but even to name. One possibility, mentioned below, would be "empirical idealism." The difficulty arises because the style of empirical idealism is both too widespread and too unknown: it pervades psychotherapeutic practice silently, calling little attention to itself, and has few spokespeople or theoreticians.

Let us imagine briefly what the mechanistic project would entail if pursued to its logical endpoint (a more rigorous engagement with the issue can be found in Chapter 2). To be consistent, the position that human plans,

feelings, thoughts and actions are a result of, say, nervous system activity would scoop away all personal responsibility, choice, and ontological status from the human subject. Logically derivable from such a view would be statements like: "I couldn't help killing him/saying that/drinking that/lying to you, because my genes, early family history and subsequent experiences forced me into it;" "I guess I still love him because I haven't left him;" "I'll just wait and see what I do: it's all mapped out for me between my brain and my environment, right?" These and a host of other questionable attitudes (including, finally, "I don't exist") follow from the reduction of the human capacities of choice, feeling, understanding, to something <u>other</u> than choice, feeling, understanding -- whether this other is conceived of as physical or mental mechanism.

The statements quoted above could be said by patients (and some of them have been said by patients) to their therapists. Few therapists, however, would respond, "Yes, you're right about that, what you say accords with my understanding of your situation" -- and yet <u>to be consistent with mechanistic theory they would have to agree</u>. Happily, most psychotherapeutic practice remains theoretically inconsistent to the extent that therapists tend to work against, rather than support, this degree of dissociation and disavowal.

As Lynne Ruder Baker has shown (Baker, 1987), a consistent physicalism, and indeed any determinist position, requires that the mental machine proceed according to content-free rules of operation. That is, one thought must be conceived of as following another not in accordance with the meaning and intentional connection of the thoughts, but from strictly mechanical causes. One brain event, whose phenomenal face is a thought or feeling, would lead to the next brain event by physical/chemical processes which can in principle be analyzed exhaustively with no reference to the thoughts they may cause.

Current philosophers of cognitive science hold a range of views with regard to the scientific status of what Baker calls "common-sense conceptions of the mental." Baker traces a continuum from such relatively accepting philosophers as J. Fodor through relatively reductive theorists (P. Churchland, S. Stich), and "instrumentalists" such as D.Dennett, who would retain intentionality and content for heuristic purposes but hold to the eliminative as the only truly scientific stance. I believe that most clinical psychologists belong in this instrumentalist camp, working contentedly with clinical entities while assuming that somewhere, somehow, a more scientific analysis of mental content could successfully explain observed phenomena in terms of mechanistic processes of some kind -- whether

the physical mechanism of brain and nervous system states or the psychological mechanism of subconscious or stimulus-response imperatives.

Clinical Theory

This dissertation arises from the conviction that the instrumentalist stance is inadequate, and that clinical psychology can benefit from an understanding engagement with the theory of mind/body relationship. If it were the case that the clinical entities of feeling, decision, thought, fantasy, are merely the outward face of mechanistic processes, then we as clinicians could indeed content ourselves with clinical experience and forget about its "real," "scientific" underpinnings. Yet both we ourselves and our patients would be operating, by our own view, in a quasi-illusory realm: acting as if thoughts and feelings were real and causative when we, at least, know they are not the root reality. We would be manipulating and seeking to investigate the phenomenal face of events whose "reality" we locate <u>elsewhere</u> -- in the unconscious, in the brain, in learning mechanisms.

Such a procedure would make of the human being we treat a kind of dupe: a causally driven entity that we support in its illusion of intentionality. By objectifying

the human subject, such a view also creates a differential or barrier between ourselves and our patients: we actually cannot regard ourselves as mechanically driven (cf. Schafer, 1976), but we can certainly regard others in that light (Ryle, 1949). Instead of lingering in the background and leaving clinical practice undisturbed, our conceptions of the scientifically justified bases of psychology necessarily influence clinical practice.

Psychology is in a special position as a science because it takes as its object entities which are also the subjects of the science: human beings. The burden and implication of this strange circumstance are more often alluded to than elucidated. The natural sciences, such as physics, chemistry, or neurology, derive a certain epistemological security from the fact that they study something other than the scientists who are conducting the study. There are many peculiarities about the position of the social sciences as opposed to the natural sciences that derive from this distinction: in the social sciences object and subject of study tend to converge. For the purposes of this dissertation, only one of these peculiarities will receive close scrutiny. This is the requirement, incumbent on psychology, that whatever it asserts or denies of human activity generally must be reasonably applicable to itself.

When psychology attempts in its experimental, clinical or theoretical mode to behave after the manner of natural science, it tends to forget this requirement. Psychology in any mode is performed by humans, and what psychology says of human behavior it says of itself. Therefore if it asserts that the human is determined, mechanically driven, reducible to physical states or subconscious causes, it simultaneously says that this, its own conclusion about human mechanism, is mechanically rather than contentually derived. Its truth value or correctness becomes mysterious on this view (who would be there to tell whether it was correct or not? To understand it? How is it distinguishable from a false view?), and a host of paradoxes emerges (see Chapter 2). As Baker has noted, psychology conceived of in this way would have no self-consistent claim to signify anything whatever, let alone to convince readers or listeners of its contents.

Clinical psychology can emerge from these difficulties only by grappling with the logical tangle at the heart of mechanism, and affirming through reasoned exploration the primary scientific status of certain mental contents. To suppose that experimental, perhaps neurological, findings will clear up the issue is misguided. As Irani has noted (1980), the most that can ever come of the correlation of brain events with mental events is just that, a temporal

correlation. No number of psychotropic drug effects, neurological/psychological correlations, or even physical healings following on mental exercise, can settle the question of the scientific status of mental events.

As we shall see, however, when mechanistic explanations are analyzed with the appropriate clarity, their contradictory nature, a kind of pragmatic paradox (Levinson, 1983), becomes apparent, and new sense can be made both of existing clinical theory and practice, as well as of new developments in the area of mind/body healing (psychoneuroimmunology). A clinical theory liberated from mechanistic bias will be able to assign the proper place and degree of importance to those aspects of mental functioning that are in fact mechanistic in flavor or physically influenced. It will also suggest new clinical practices based on an understanding of the free human capacities, and recognize as a real entity the responsible human subject.

The Multiplicity of MindBody Problems

There are a number of issues in mind/body theory that have persisted, despite some evolution, from the time of the Renaissance. In this dissertation, one such issue has been raised for detailed discussion because I believe it has the clearest possible resolution and the widest possible

clinical implications. Before resuming the narrow focus of the discussion to follow, I want to review several common mentalist claims and their mechanist rebuttals.

One mentalist claim, put forward by T. Nagel in his celebrated essay, "What is it like to be a bat?" (1974), involves the irreducibility of mental or sensory qualia, the feel of their direct experience. We cannot avoid the impression that, though we have never felt it ourselves, there is an irreducible somewhat to the experience others have of color, their own feelings, or (if they are bats) the sensations of their flight. Our own sensations, inner and outer, have if anything a stronger claim to irreducible, felt quality, and it is hard to see how this mental reality could be explained in purely physical terms. This line of argument dates back at least to Leibniz (Monadology, 1714). Leibniz pointed out that even if a thinking, feeling machine were somehow constructed, and were large enough for us to walk around in and observe, we should see there only "pieces pushing against one another, but never anything by which to explain a perception." The rebuttal, as put forward for example by P. Churchland (1988), recalls that we now have a comprehensive understanding of the light frequencies that produce sensations of different colors and we may one day have similar data for all qualitative experience. It

further notes that machines operating in purely physical ways now perform calculations once thought to be the exclusive domain of human intelligence. Finally, it tends to spend a little of the infinite future on the problem: we have made great strides; we have only begun to investigate these matters; we have all of scientific futurity to work away at it; surely in time the felt qualia of experience will be describable in purely physical terms.

Another favorite mentalist or anti-reductive (for example, anti-artificial intelligence) claim is that no physical or mechanistic process can account for human creativity. Whereas humans continually come up with new insights, solutions and creations, no machine ever could, and the human achievements of this kind accordingly cannot be explained mechanically. This view has often been put forward by humanists feeling a threat in the growth of materialist experimentation and technology. at least since the time of G. Vico (1668-1744). The privileged place of human creativity has been opposed by, for example, M. Minsky (1985) who cites the invention of computers that have been taught to "improvise" agreeable jazz music, play "inspired" chess, and otherwise mimic the supposedly free range of higher human functionings.

Still a further mentalist position emphasizes the self-aware or self-controlling feature of human

consciousness as distinctive, irreducible, and impossible to account for in mechanical terms. In a sense, the Cartesian cogito (1637) inaugurates this position for the modern world. As we shall see, there are specific problems with the cogito, and its use in Cartesian dualism obscures some of its continuing relevance. As we shall also see, it was partially anticipated in a pictorial fashion by one aspect of Aristotle's psychology. Typical replies from materialist quarters would include A. Turing's famous (1950) article on "Computing Machinery and Intelligence," which asserts that self-referential functions are certainly both conceivable and available in the machine world, and that human consciousness has no distinctive status in this regard either. His position has received a considerable boost in recent years through the advent of Parallel Distributed Processing and other computer and mechanical achievements (McClelland et al., 1986). We shall examine some of the fundamental issues in the artificial intelligence debate in the second and third chapters of this thesis.

Empirical Idealism

A proper perspective on these relatively contemporary topics in the mind/body debate can be gained through the analysis of pragmatic paradox (called "cognitive suicide" by

Baker) that appears in my second chapter. It has a number of advantages, as a focus for discussion, over the formulations mentioned above.

First, it is directed to the current moment of discussion, rather than to some future, past, distant or hypothetical performance. That is, where other formulations of the mind/body question look at relatively unavailable, even totally hypothetical data (e.g. the Turing test), this discussion focuses on the immediate circumstances of its own production. We need not wait for the discovery of physical correlates to our behavior, or for advances in computer technology. We have immediately available at least one point for investigation, namely the data of our current theorizing and discussion. What can we see about the conditions of human psychology based on this data?

Second, the focus on pragmatic paradox in Chapter 2 examines in effect the qualia of the mechanist rather than the qualia of the mentalist, and asks if the mechanist's own qualia are being consistently disregarded. The strong mechanist position requires that all mental phenomena be ultimately capable of explanation in a formalizeable framework and in terms of something other than conscious mental events. Do current mechanisms consistently treat their own statements as if this characterization applied to them?

Finally, the focus on pragmatic paradox unites in one gesture, as we would expect from a theoretical approach that surveys both sides of a duality, the physicalist urge for empiricism and the mentalist urge for understanding. That is, the investigation at hand, though "theoretical," by no means abdicates the claim to cogency deriving from empirical studies. To encounter the limits and conditions of our own current thinking is to encounter empirical data of the most compelling kind.

These points amount to an affirmation of the alternative to psychological mechanism, namely empirical idealism. It can be contrasted with empirical realism as represented by physical neuroscience. When neuropsychologists refer to the events of the synapse, to neurotransmitters and to other sites of electrochemical activity, they are always referring to distant and tentative experimental results. These results can be confirmed or disconfirmed only by very few specialists in highly specialized circumstances. They are elaborate reasoned derivations from the actual observed processes in each case -- generalizations and extrapolations from observed experimental data. Their ultimate explanatory pretensions are based on further extrapolation from current achievements in such fields as medicine and neuroscience to a fantasied

future in which machines do all that humans do or in which human behavior is strictly predictable (Pagels, 1988). In sharp contrast to this, empirical idealism notes that we each have continual access to our own understandings; in fact, we never come beyond them to some more fundamental realm. They are truly empirical data. When, therefore, this thesis encounters the pragmatic paradox of mechanistic theories, and understands thereby the irreducible reality of (some) mental content, this is no mere "theoretical" conclusion, but rather an experimental one.

The rhetorical power of science derives largely from its variations on the theme of unmasking: "What seems to the layman like X is really (made up of) Y." Every such Y, in turn, suffers reduction to some other process, Y'. This strategy of unmasking loses all force in one special case, however, which is the moment of seeming itself. Here no reduction is possible, and empirical idealism comes into its own. The current moment of seeming -- the moment of evidence or of understanding -- always receives our naive realistic faith. This is the last Y in the chain of unmaskings, and cannot be consistently reduced.

One reason for the narrow focus of the present dissertation lies in the desire to avoid an error that might

be called the Cartesian Slide. Anyone reading the Discourse on Method must be struck by the combination of rigor and laxity that appears in the course of the famous fourth chapter with its cogito. Descartes embarks on a project of universal doubt, but once this doubt has convinced him of his own undoubtability all rigor is lost and every conclusion Descartes desires seems to him to be justified logically. Contemporary readers who follow him through the undoubtability of doubting itself will shrink back at his very next argument for an equally undoubtable divinity. In the present dissertation, which seeks to perform a related exercise in thought to that of Descartes' radical doubting, the attempt is made to avoid the Cartesian Slide into agreeable consequences, and to restrict the inquiry to its moments of greatest cogency.

Chapter 1: Historical Perspective

The terms of current debates on the mind/body problem derive from the long history of the issue within psychology and philosophy. By and large, current positions reframe Renaissance and Enlightenment theories. Therefore historical reviews of the problem often begin with Descartes, whose dualistic formulation inaugurated much of the subsequent debate. It also provided the theoretical basis for materialism not only in philosophy but in its technological applications, and Descartes explicitly intended his system to justify the investigation and manipulation of the material world for human ends. To see the current positions in perspective, however, and apply this perspective to clinical practice, we must go back briefly to examine the origins of the question in ancient Greek philosophy. This procedure presents us with an evolution of mind/body problematization: the Greeks did not pose the same questions we do, nor even as the Enlightenment did, and this difference is instructive. It is also

relevant to the theme of this dissertation, since Aristotle's tripartite view of the human organism offers a possible alternative to Cartesian dualism.

After a brief review of the relevance of Classical approaches, I will skip to such later incarnations of the issue as have a bearing on my themes of mechanism and reflexivity. This chapter therefore does not aim at recounting anything like a history of the mind/body problem, but rather grounds my discussion in history and exposes by contrast some of the limitations of present-day reductivist schemas.

Classical Approaches

Taking a middle position between the Platonic concentration on pure Forms and Democritus's reduction of all things to atoms, Aristotle conceived of the psyche as a life-giving and life-filled entity. He was a physiological psychologist (the first) in the sense of trying to localize mental function, yet he also saw the "nous" or mind as untouched by bodily events, including death. Robinson's (1986) review of Aristotle's psychology suggests that Aristotle held a tripartite view of man, with body and psyche interpenetrating during life and dissolving at death, while nous (mind) existed eternally, accessible as a

capacity to humans during their lives. Thus, Aristotle derived such functions as sleeping, emotion, memory, dreaming and learning from the senses and other biological bases: and these functions he assigned to that part of the mind he called psyche. The higher mental functions of combinatory reason and perception of universal forms were assigned to nous and epistemikon.

Robinson therefore finds a dualism in this tripartite system; Aristotle's distinction between the "moveable" body, with its attached psyche, and the "unmoveable" epistemikon or capacity for perceiving universals and spirit. The soul (psyche) and body are one kind of entity, while the epistemikon is, or at least opens man to, a different kind of entity -- the Forms, the "unmoveable," and higher reasoning.

Yet the distinction between moveable and unmoveable, like other dualities to be found in the ancients, should not be too readily equated with any modern counterparts. For one thing, this epistemikon that perceives forms is called a "part of the soul," rather than being outside or in opposition to it. Erik Ostenfeld reminds us (Ostenfeld, 1986) that looking for contrasts in Plato and Aristotle to parallel Cartesian mind/body dualism is a hopeless task. Aristotle in De Anima clearly derives body from soul: that

is, the soul makes the body rather than the other way around. It pre-exists as the "form" of the body, and form for Aristotle as for Plato is accorded a higher reality status, a higher degree of "actuality," than is accorded to matter or body. When Aristotle defines the soul in De Anima, in fact, it is in just this way: as the original form of the body. He uses an analogy: The soul is to the body as axe-ness is to the axe.

Ostenfeld summarizes a part of the distinct flavor of Plato and Aristotle's psychologies by diagramming their un-Cartesian division of the parts of the human being. Where modern thinkers are likely to group life with mindless matter, Plato groups it with thought and consciousness. The Greek word "psyche" in fact is translated both as "mind" and as "life" in English. For us as post-Cartesians, the concept "body" is more likely to include the organism's life, while for the ancients life and thought were grouped as distinct from body. As Ostenfeld says: "For Plato the soul is the principal of life: when it leaves the body, bodily life comes to an end. For Descartes the converse is true: when the mechanical, autonomous body-machine stops, then the soul leaves it." (Ostenfeld, p.28).

Moreover, Aristotle and Plato agree that all the world of things, as well as humans, is preceeded by the

intelligible forms of things. Their renowned disagreement about the instantiation of ideas, and the importance or irrelevance of focusing on sense perceptible reality, should not obscure this priority of the world of idea and potential to the world of fact and example. Aristotle, if anything the more empirical and physicalist of the two, goes so far as to say that the soul is all things (De Anima, Chapter 8).

To summarize these observations, we could say that for Plato and Aristotle the mind/body relation was a topic and a problem, but its terms revealed an outlook hard for us to recapture today. In some sense, they regarded all things as ensouled and alive, and life and mind had an intimate relationship apart from their association with the physical stuff of bodies. For the Aristotle of De Anima, a tripartite human organization included body, psyche and nous, in which psyche corresponds to what we could call life/soul and nous corresponds to what we might now call spirit or intellect.

The Role of the Epistemikon

Aristotle's definition of soul as the form of the body suggests another difference between his views and those of either Cartesian dualists or materialist monists of our time. Descartes saw mental and physical laws as quite

distinct, and in fact his innovation was to make them so. Materialist monists and epiphenomenalists (e.g. Churchland, 1984) see the mind as non-existent, but if anything they would regard its phenomena as the result rather than as the cause of bodily processes. For Aristotle, however, the soul was the life of the body, regulating its movements and maintaining its form.

This is hardly our lived experience today, any more than it is the prevailing view of the medical and psychological professions. We can introspect on our thoughts and feelings, and we can sense bodily events, but we hardly feel our "minds," our soul-stuff, to be the principle of organization by which the body is directed. The mysterious connection between our feelings and thoughts on the one hand, and bodily events on the other, is normally experienced principally through the phenomenon of willed bodily movements. In the Aristotelian system, however, the relation of psyche to body is analogous to that of mold to sculpture or blueprint to building, though these examples omit the quality of life, of movement-within-stasis, that informs his view.

The special care Aristotle devotes to the epistemikon, that within us capable of conceiving the universal Forms of things, merits further attention and

represents a crucial contrast to the rest of psyche. Since the epistemikon is open to receive the universal Forms, says Aristotle, it must itself be un-formed -- open and infinitely malleable. In this way, it can take on the form of the Form it receives at any given moment. This quality of the epistemikon, explicitly emphasized by Aristotle, is at the heart of his psychology. Whereas psyche has a (fluid but stable) shape, a set configuration and pattern of operation, and makes the body according to this pattern, the epistemikon has no shape, no fixed pattern, but is essentially, radically un-formed. In its un-formedness, it represents a break within the closed system of psyche and body. And it is within this system: recall that the epistemikon is defined as a part of the psyche. It is a kernel of un-formedness, therefore, within the established pattern of the mind/body unity.

The relevance of these considerations to a discussion of mind/body healing should be obvious. All disease, mental and physical, represents a set pattern, a fixed constellation of signs and symptoms (Shapiro, 1965, 1981; Hoffman, 1981). If it were not so, we would not recognize the recurrence of a disease from one person to the next, nor even from one day to the next in a given person. The tragedy of all disease forms is their over-stability, their

fixedness (even if it is the fixed quality of hyper-lability, as in hysteria). This has been recognized in systems theory and family therapy (Hoffman, 1981), in which a pathological homeostasis has come to be regarded as a common element to dysfunctional states of all kinds.

Although Aristotle's concern is not specifically with pathology, I propose that his model can be extended to cover the Formal causes and the bodily manifestations of disease as well as health. The healthy or unhealthy psyche would then be the determining form of a healthy or unhealthy body; its homeostasis would correspond to a homeostatically maintained body state. Yet if this were the whole story, there would be no avenue for healing, no chink in the determined and lawful system in which psyche and body are established. The presence of a part of the soul that is open to the Forms, however, the presence of the epistemikon, guarantees that the mind/body homeostasis contains a healing flaw. By virtue of its unformedness, its radical openness to new and higher influences, this free part of the soul remains unafflicted by any diseased homeostasis. It represents the potential for change within the mind, and therefore, since the mind is the life of the body, within the bodily organism as well.

Aristotle begins De Anima with a reference to the peculiar circumstance that in discussions of the soul it is the soul, psyche, that knows itself, discusses itself. We might ask what part or faculty within the soul has this ability to to know and examine the very soul to which it belongs. Since we have Aristotle's definition of psyche as the Form of the body, and since the epistemikon is the only part of psycne that can know Forms, it follows that psyche knows itself through its epistemikon. This perhaps obscure formula has important consequences. To rephrase it, we could say that it requires unbounded, unfixed energies of the mind for the mind to know itself. Only that within the mind which is pattern-free, unfixed and open, has the capacity to know the mind's own patterns. To reverse the emphasis: the emotions, perceptions and other events of soul-life are investigated in psychology, and this indicates the existence of something in the soul that is free from any fixed patterning. Goedel's famous proof, as we shall see in the next chapter, points in the same direction, away from reductionism and toward the capacity for human change through self-investigation.

Substance

Formulations of the mind/body relation from the time of ancient Greek philosophy through the Renaissance returned

over and over to the question of substance. In Aristotle, as we have seen, reference is made to one aspect of substance, that of the Forms, which precedes and gives rise to another order of substance, that of the body. He distinguished these as secondary and primary substance, respectively. Although in Plato the question of the substantiality of soul is answered variously and ambiguously (Ostenfeld, 1986), in Aristotle the soul is more definitely a substance of a different order than the body and distinct from it. The distinguishing mark of primary substances, such as the individual human body or other sense-perceptible object, is their capacity for self-subsistence: they exist by themselves and apart from derivative accidental qualities such as the fact of their perception by different observers (Vesey, 1964). The distinguishing mark of the secondary substances, such as the Forms, is their eternality and unmoveability.

In Descartes, too, distinctions among substances frame the discussion, but with a difference. Vesey shows how Descartes proceeds from the definition of substance appropriate only to Aristotle's primary qualities when he calls a substance "a thing existing in such a manner that it has need of no other thing in order to exist." (Vesey, p.12) In his Principles of Philosophy (1641), Descartes used this

definition of substance to argue for the distinction between extended things and thinking things as the fundamental duality in the created world. Since extension could not be thought of apart from matter, nor matter apart from extension, and yet both could be conceived without reference to thinking, he concluded that extended corporeality (res extensa) was a separate substance from any thinking thing (res cogitans). This, in turn, could be seen as self-sufficient and needing no extension, which is foreign to it, in order to exist. Both substances are grounded for Descartes in the original substance of God, the only truly self-subsisting essence. The criterion of different substances here is simply one's ability to think of the substances clearly and distinctly; this becomes the sole guarantor of their independence from one another:

> And although ... I possess a body with which I am very intimately conjoined, yet because, on the one side, I have a clear and distinct idea of myself inasmuch as I am only a thinking and unextended thing, and as, on the other, I possess a distinct idea of body, inasmuch as it is only an extended and unthinking thing, it is certain that this I (that is to say, my soul by which I am what I am), is entirely and absolutely distinct from my body, and can exist without it.
> (quoted in Vesey, p.29).

It is important for our purposes to note the dramatic shift from Aristotle's conception of substance that has gone on here. Where Aristotle's distinctions were among orders of

substances, primary and secondary, and derived individual existences from their eternal essences, Descartes' formula reduces mental and bodily events to an equal status. Res extensa does not derive from res cogitans as the Aristotelian body had derived from its form, the soul. Rather, both have a sibling status and derive from God. This separation of mind from matter, and the identification of the latter with the body and the body's life, constitute the re-ordering of substances that made Descartes so revolutionary.

Since Descartes distinguished these sibling substances by means of a criterion (their seeming independence from one another) that Aristotle had applied only to one level of reality (e.g. primary substance), Descartes' system in effect turns all of "mind" into a similar bodily substance with the physical body, and so begs the question of their relationship. Descartes' celebrated act of distancing, of setting up a gulf between mind and matter, is thus revealed to be secondary to his act of equalizing them. He was in fact uniting, at one level of reality, categories that he then distinguished on that same level.

To reiterate: Descartes's achievement, which determined subsequent formulations of the problem, was

precisely the shift from a hierarchical focus on orders of substance to a philosophically egalitarian focus on kinds of substance. By this means, he created a commonality between mind and matter that separated them more effectively than previous conceptions could ever have done. The "problem" arose as to how body and mind could interact -- how an extended thing, which occupies space, could be affected by a thinking thing, which does not. My earlier review of Aristotle's model suggests that this problem is an artifact, a Cartesian invention. The problem did not (yet) exist for Aristotle because body and psyche were seen as united but distinct levels: psyche was the Form of the body, in-forming it from above. The Aristotelian "moveablility" of both psyche and body, as contrasted to the "immoveability" of nous, suggests that if distinctions of kind were to have been made, they would have left psyche and body together in one category, yet with psyche as a formal or ideal pattern that gave bodily organization its life. In other words, the principle of hierarchy could account for a relationship that the principle of typology only later was to make problematic.

Post-Cartesian philosophers who approached the problem, however much they disagreed with Descartes, accepted in one way or another, not the exact substantial

distinction he made, nor even the focus on substances at all, but rather the problematization in terms of kinds rather than orders.

My earlier inference that, of all parts of the Aristotelian human organism, it must be the epistemikon that can know psyche from the inside demonstrates, by way of contrast, the significance of this shift. In Descartes, there is no principle corresponding to the epistemikon, no radically unformed organ of perception within the soul that can be used to perceive the soul itself. The Cartesian substances have the effect of obscuring the necessary existence of something unconditioned, unpatterned, if we are to somehow notice anything whatever about the soul and the body -- including all patterns and all conditioned responses. It obscures, that is, the existence of a higher order of cognitive functioning. From Aristotle we derived the formula that it takes an unformed portion of the soul to make pronouncements about the soul. When the body and mind are sharply distinguished, however, and awareness of some principle of unformedness such as the epistemikon is lost, then we are left with mechanical nature on the one hand and an unrelated, equally mechanical soul on the other. Such a soul would have no clear influence on the body a) because it

is a different kind of substance; b) because it does not exist as clearly (measurably, visibly) as the body; and c) because it is itself mechanically determined.

We can view the transvaluation of what are now called mental and physical events as the rotation of an axis with the category "mental" at one pole and the category "physical" at the other. In Classical times, eternal Forms and higher reasoning were located above, and worked into the physical bodies that reflected them below. The axis had rotated ninety degrees by Descartes' time, so that mental and physical were on an ontological level. Instead of Forms causing bodies, bodies and Forms are side by side, defined as different kinds of substance, both caused by God. This move prepared the way for a later physiologically-influenced philosophy to continue rotating the axis through another ninety degrees, so that physical/bodily events are now, in the wake of logical positivism, widely thought to cause mental ones. A final step in this long descent occurs when brain events are considered the only real processes, and mental events are considered either epiphenomenal or illusory (Churchland, 1984; Minsky, 1985).

Descartes' famous cogito, with its mentalistic emphasis, should not mislead us. For all that the cogito demonstrated his mentalism and his conviction that there is

a reality to the individual self, his whole project had the effect of elevating matter to a new philosophical respectability. Its later, further elevation to the sole explanatory principle of the modern sciences derives from the Cartesian split, which I have described in terms of distinctions among kinds, rather than orders, of substance. The post-Cartesian, often anti-Cartesian views of Leibniz (parallelism), Spinoza (the dual-aspect theory), La Mettrie (radical physicalism) all accepted the underlying tendency away from an understanding of orders in the Aristotelian sense. As K. D. Irani has shown (1980), the descendants of these views in our time, including logical positivism and behaviorism with their principled refusal to consider mind/body interactions, have derived in a complex fashion from this same development. Attempts to define the problem of mind/body interaction out of existence (Ryle, 1949; Ayer, 1950; Wittgenstein, 1953) either implicitly grant a primacy to physical matter or at any rate demote mental events from a substantial status.

The Concept of Mind

In our own time, these trends, while exceedingly various, have received new impetus from developments in computer science (Rumelhart et al., 1987), from

psychologically related physical cures (Justice, 1988), and from neuroscience (Kolb, 1985).

Although the trend towards materialist reduction (as in the cheerful quotation from Wortman in my Introduction) has prompted much theoretical ferment, especially among cognitive scientists, there are few mechanists who engage with the problem of self-reference as directly as Ryle in The Concept of Mind (1949). Ryle's project is to eliminate the concept of "mind" from scientific discussion, and to recognize in it a superstitious holdover from outmoded, undigested philosophical prejudices. He particularly regrets the Cartesian opposition of a mental world to an extended physical world, in part because of the implication that nouns and predicates referring to the former will invalidly borrow status fromt the latter. The rose's shadow, as Ryle might say, ought not to boast of its scent.

In a typically charming passage, Ryle attempts to demystify the apparent authority of the subject and the confusing nature of the word "I":

> An ordinary reviewer may review a book, while a second order reviewer criticizes reviews of the book. But the second order review is not a criticism of itself. It can only be criticized in a further third review. Given complete editorial patience, any review of any order could be published, though at no stage would all the reviews have received critical notices..... This, I think, explains the feeling that my last year's self, or my yesterday's self, could in principle be

> exhaustively described and accounted for by me, but that my today's self perpetually slips out of any hold of it that I try to take. (Ryle, 1949, p.196)

In this analogy, the "I" corresponds to the most current book review. It cannot contemplate or discuss its own current moment of reference, but only a past moment, however instantaneously recent. What "I" choose to do or say looks utterly free and unpredictable when seen in this light.

Ryle is at pains to explain away this problem, which he calls the "systematic elusiveness of 'I'." By analyzing it as an essentially temporal peculiarity of first-person utterance, he means to show that the "I" is no different in kind from any third or second person activity. He concludes that since we feel entitled to declare other people's acts and futures to be in principle predictable, we must resign ourselves to our own predictability and explicability. It is merely our self-proximity that makes things appear otherwise.

Ryle's move here involves an appeal to democratic principles, as if the special properties of first person speech ought to be dismissed on grounds of unclubbability. Indeed, if we are really content with the predictability of others, we should in good fellowship confess to our own. But to assume this predictability of others without establishing it is a <u>petitio principii</u>, since our own and

others' mechanistic explicability is precisely the unknown and mysterious question brought up by Ryle's reflections on the use of the word "I."

As Chapter 2 will discuss in greater detail, the decisive moment for such discussions is always the moment of current utterance. When writing his chapter on the systematic elusiveness of the "I," does Ryle consistently behave as if he really believed his insights to be predictable? For if they are predictable, then they did not come about by insight into the state of affairs being discussed. Insight or understanding cannot be predicted, since they would then be determined processes. They would be "blind": <u>having</u> to come about, they would not arise in response to the object of understanding, but in reaction to some precipitating cause.

Rather than arguing from a predictability assumed to apply to others and extending this condition to our own acts and insights, we must therefore proceed in the opposite direction: we first acknowledge the unavoidable, inescapable burden of insight in our own case, and then extend it to our neighbors. As we shall see, it is only this direction of extrapolation that can do justice to our inability truly to divest our own subjectivity of its instantaneous self-trust.

Reflexivity Theory

In our time, questions touching on reflexivity are widespread in psychology, literary theory, and philosophy. The existing literature raises and wrestles with such themes, perhaps without the urgency that psychotherapy gives to theoretical questions. Hilary Lawson (1985) has grouped together Nietzsche, Heidegger and Derrida (the selection could be extended) as theorists whose writings bear importantly on self-reference. Lawson goes so far as to call reflexivity "the post-modern predicament" in his subtitle. We cannot now ignore the paradoxes inherent in self-referential language and thought, says Lawson, and it is our duty to incorporate these paradoxes, at least in the form of some kind of constructivism, in our understanding of life and literature. His work attempts to both cite and exemplify questions about the capacity of language to refer to anything outside itself. In various ways, Nietzsche, Heidegger and Derrida are shown to embrace logical and linguistic contradictions that a pedantic rationalism would forbid.

While the works of these theorists (see also Wittgenstein, 1953; de Man, 1971) contribute involutions enriching the field, the present thesis meets them only at

the point of their clinical relevance. This relevance derives from the circumstance that the writers in question are or were real people, whose work was motivated and intended -- as with all human acts. The recent spate of publications surrounding Paul de Man's anti-Semitic writings during World War II has put to the test any radical questioning of language's referential properties. This recent academic scandal has relevance to our theme precisely because it breaks down the barrier -- all too solid in scientific psychology -- between object and subject of study.

Without, for example, drawing (and then necessarily shirking evaluation of) the contentual relationship between de Man's early anti-Semitic writings and his later deconstructions, we can still note the disparity between deconstructionist questionings of referentiality and the ferocious historical realism of de Man's posthumous defenders (Simpson, 1989). Very simply, <u>among friends</u> we find that such philosophically disputed items as personal identity, linguistic reference, and conscious mental events, are all taken in naive realistic fashion as self-evident reality. We automatically set a limit to our doubt as to the reality of mental life, the power of language to communicate, and our own existence as subjects, just beyond whatever is personally important to us.

To say "the brain thinks" (or any analogous formula) is to make of psychology another forum for paradox. For it is we -- the brain-bound ones -- who make the physicalist or mechanist claim. And when making it we do not really believe its truth or correctness to be, for example, chemically determined. Thus, the "limits" mechanisms assign to language and thought have limits of their own, beyond which lie unexpected expanses. If we manage to see through the claims of mechanistic psychologies, however, we are left with the uncomfortable burden of personal freedom and responsibility (Pagels, 1988). So irksome is this burden that we must wonder whether mechanistic psychology did not arise and flourish with the unavowed aim of avoiding it.[1]

Chapter 2: Reflexivity

As we have seen, the tendency of modern psychology and medicine has been to accept as given a physically extended world of material objects, and to derive mental processes from this world. As Irani has pointed out (1980), reduction does not always mean that that which is reduced to or derived from another realm does not exist -- only that its origins or causes or reasons have been clarified. Thus, there are those who do not deny there are instrumental uses and meanings to the mental, yet the guarantor of scientific validity remains the physically observable (countable or measurable) processes. There exists what might be called a gold standard of physicality: we can play around with mental terms, discuss emotional and mental and willed events, yet we know that ultimately it is physical processes that are somehow underneath or behind these events. The physical processes, or mechanistic processes modeled after those observed in the natural sciences, "back" our discussions of mental events as the gold standard was once thought to

guarantee the validity of a currency. Even if no direct appeal is made to physicality, the mechanistic nature of psychological theories is a lineal descendant of physical law.

There are reasons for this bias in psychology and medicine, some of which were discussed in the previous chapter. Besides the more narrowly historical and cultural reasons, we note the availibility and confirmability of the objects of experimental method. Whatever behaviors can be counted, measured or weighed after the manner of physical objects can also be counted, measured or weighed by subsequent researchers. An impulse toward objectivity -- in the sense that scientific results should not be influenced by the personal predilections of the scientist -- stands behind this experimentalist tradition. The present chapter intends to show, however, that in the field of psychology the gold standard of physicalism distorts our view of the psyche. It can be discarded without abandoning the principle of scientific objectivity, though not perhaps without modifying how this principle is understood.

When cognitive psychology likens mental events to computer programs (Johnson-Laird, 1988), when psychoanalysis invokes biological or deterministic underpinnings (Shapiro, 1965, 1981), when behaviorism locates stimulus-response

interactions at the base of our acts (Schorr, 1984), then we have a psychological theoretical landscape modeled on physical science that leaves no room for the un-determined, non-physical human being as the source of his own potential life. All these views regard man as the product of finished, fixed processes, as if pscyhology were a subset of biology -- while biology in turn is ultimately based on the model of mechanical physics (Brady, 1981). This is even, or especially, true of the newer "chaotic" sciences which include such lifelike and mindlike properties as non-periodicity (Gleick, 1987). Such views, if pursued rigorously, would all require the theoretically consistent clinician not to speak to the patient as to a relatively free, relatively insightful person (with the clinician trusting that the patient has these qualities, just as he cannot fail to trust them in his own cognitive life), but instead to address the patient as one bundle of determined processes interacting with another. Of course, we are not always true to our theoretical bases, and this is clinical psychology's saving grace.

When I confront another human being, what do I have before me? Why would I address this form with words? If the intention is to help, what do I imagine are the effects of my words? Taking the person as a physical entity, or a

mechanistic formation of any kind, I can only hope to speak words or administer other treatments that will interact in a predictable way with the processes in that entity so as to change undesired behaviors into desired behaviors. As a cognitive behaviorist I might speak so as to change the individual's "programs;" as a psychoanalyst I might speak so as to change the expression or configuration of unconscious drive states; as a psychiatrist I might administer drugs that have a certain effect on biochemical processes in the brain. All these modes of treatment take the person as a mechanism, and faithfully treat him as such.

Yet psychotherapeutic treatment does not go on in anything like this fashion. Rather, we address patients as if we assumed that they have the same (limited) freedom in cognition and action as we do ourselves, and we address them as creatures who like us live according to meanings. Our theory should reflect this procedure and explore its implications, rather than make appeal to a physicalist backing, the "gold standard" mentioned above, according to which the human being is understood as operating by known physical or psychological laws which can be, or at least in principle could be, delineated in their entirety. The central problem with this imaginary scientistic program is its paradoxical, self-defeating nature. In this chapter we

will examine the paradoxes inherent in theories that regard the human being as a mechanistic or natural process.

Reflexivity, or the peculiar capacity of humans to make pronouncements about their own behavior, has generally unnoticed implications for the theory and practice of psychotherapy. In particular, the fact that it is we ourselves, and no one else, who discuss and describe what humans are like makes certain kinds of pronouncements invalid. Self-reference has both logically permissible, and logically impermissible, in fact senseless, forms. One aim of the current chapter is to distinguish between these two forms. As we shall see in the final chapter, this logical point has consequences for treatment. If we focus here on its logical aspect, it is to suggest that the clinical notions of "rapport," "spontaneity," "growth," and related concepts have as sound a theoretical base as any concepts in science.

Where botany can declare, "In this region there are no flowers," psychology cannot declare, "Here there are no people." This is the case because it is not some abstract "Psychology" that would make this claim, but an actual person. Psychology, if it talks about humans in general, necessarily talks about psychologists as much as about

anyone else. Whatever it asserts of all humans, as we stated in the Introduction, must be capable of extension to psychologists themselves, especially at the moment of their theorizing, for that too is a human moment. Therefore special burdens are incumbent on the theory of psychology, as opposed to the theory of other sciences. It cannot disregard, as other sciences can (at least, to a greater extent), the circumstances of its own proclamation.

We can hardly say, without clear contradiction, "There are no words." The problem with such a sentence is that it makes use of the very elements it denies. It could be called a joke, or a lie, or a logical paradox, or simply meaningless. (Just how we would categorize it would in fact depend on our understanding of the speaker's motive in making the statement.) In exactly the same way, we cannot claim our auditor's assent to statements such as "I do not exist." Most of us would agree these sentences invalidate themselves in some fashion. Yet we tend to overlook the similarity to such paradoxes in the widely held mechanistic viewpoints that maintain there are ultimately deterministic, probabilistic, physical, or computerlike processes at the basis of human psychology. In other words, statements such as "the brain thinks," or "your mind operates by program" must also be seen to invalidate themselves if psychology is

to have a consistent theoretical foundation. They too are logical paradoxes, for their very assertion indicates the speaker believes in unforced insight into correct understanding -- a kind of insight that the content of the statements would attempt to obviate, reduce or exclude.

I suspect that a really healthy response to the statement "the mind is a result of brain activity" would be hearty laughter: its absurdity should be immediately apparent. It is not apparent, however, certainly not for the mental health field as a whole, and so we shall approach the matter from a variety of angles to look more closely at the problem with such formulations.

This investigation represents the heart of the present thesis, and its conclusions are meant to provide the foundation for a view of psychotherapy and psychoneuroimmunology that gives due credit to the thinking, feeling and willing human being as a real entity, not derivable from or reducible to any <u>other</u> processes understood to be more real or to have more explanatory power.

<u>The Physicalists' Oath of 1845</u>

The philosopher Hans Jonas put the issue succinctly when he referred to the famous Physicalists' Oath of 1845

(Jonas, 1981). I will examine Jonas' argument in some detail, supplementing it at times, because his point with regard to the Physicalists' Oath has important ramifications. This oath was sworn by a number of young physiologists in Berlin, first du Bois-Reymond and Ernst Bruecke, then Helmholtz. They vowed to explain the human organism exclusively in physical-chemical terms. For them, this mechanistic view of man was part of a struggle against the prevailing vitalist schools of the day, but it has since become so much the accepted orthodoxy as to scarcely need to be mentioned in medical and psychological textbooks.

Jonas asks the crucial question of why these ambitious young scientists, if they really believed in the chemical-mechanical thesis to which they pledged fealty, would have the need to swear an oath to secure their future actions:

> All three, each about to ascend to great fame, remained true to the goal of their youth -- with brilliant scientific success. What they did not realize was the fact that by undertaking this promise (for it was their "oath"), they were already behaving contrary to the peculiar contents of <u>this</u> promise.... By the fact of the oath they were imputing to something altogether non-physical, to their own relation to the truth, just that power over the behavior of their brains that in the content of the oath they were denying for the general case. (Jonas, pp.13-14)

Jonas goes on to suggest that the physiologists' ruling-out of mental life from an explanation of the human being was

correct, insofar as the sense-perceptible aspects of the human being can be coherently explained up to a certain point in physical terms alone. Du Bois-Reymond was further correct, on this view, in his later "ignorabimus:" the declaration that we shall never know the sources of mental life. That is, he was correct to the extent that a viewpoint which remains within physical laws will never account for what Jonas calls subjectivity. Despite this limited appropriateness to their stated goals, the strong determinist, physicalist stance of the Berlin group undoes itself logically, as we have begun to see. By swearing loyalty, by banding together for a common ideal, even by saying that the human can be explained in mechanical, chemical, physical terms alone, these men were also making the opposite claim: that it was their wills, their beliefs, their goals that were sovereign, not their brains and metabolisms. Else why bother to swear an oath? Or did they swear it, asks Jonas, only because the physiological processes governing their every thought and action prompted them to do so?

The Berlin physiologists and their intellectual descendants, if pressed, would probably answer the latter question in the affirmative. The correctness, even the meaning, of their commitment to physicalism becomes

problematic with this affirmation, however. If it had to be sworn due to physical/chemical causes, then it was not sworn because of its meaning, or content. It suffices to ask if at the instant of swearing the oath they believed it to be a matter of physio-chemical necessity that they swear it.

Strong Views and Negative Theses

The structure of Jonas's critique of the Physicalists' Oath involves a negative comment on a strong view. That is, he finds an instance of extreme physicalism and declares its paradoxicality. The thesis is negative, in the sense that it negates physicalism rather than proposing and investigating some other explanatory principle. The physiologists' view is "strong," in the sense that it pretends to total explanatory power over the human organism based on mechanical-chemical laws.

Today there is a rich array of positions on such matters as mental determinism, the degree of independence of thought from brain states, and related problems (see Gardner, 1985). Only in the confrontation with a strong view of physical determinism can the paradoxical nature of mechanical explanations of the human being be seen for what it is. This paradox, once clearly understood, then has implications for all theoretical admixtures and derivatives.

The subjectivist thesis of Jonas has what I have called the negative form of decrying a contradiction. It is an instance of reductio ad absurdum, that is, an argument that begins by assuming the premises one expects to disprove, and shows that they lead to absurd consequences. To some degree, the need for this particular style of discussion derives from the history of mind/body problematization. For the ancient Greeks, certain base-points of discussion were included in the realm of the obvious that are not at all obvious today, and logical argument proceeded comfortably from these very different premises, as we see in Plato's dialogues. For example, it was not considered problematic to refer to the soul, or to its immortality, as real processes. In our own time, it is the quite available realm of the subjective -- inwardly perceived thoughts, feelings, perceptions, mental images -- that is the questionable element. For Jonas to simply declare it as real would hardly be adequate: its ontological status is precisely the point at issue.

Instead, Jonas confronts the strong view of physicalism -- including in it artificial intelligence theory, systems theory, biochemistry and neuropsychology in their strong forms -- and asks after the implications of its very assertion, of the fact that someone bothers to announce

it. Just as with the physiologists he focused on their oath, without denying their scientific achievements, so he implies that more current physicalisms and mechanistic explanations contain similar hidden beliefs in that which they purport to deny. As with the physiologists' oath, this hidden belief is found in use, in the paradoxical stance of reductivist writings. In psychology, we can look at the statements and promises, the beliefs and viewpoints, in Jonas' term the subjectivity, that goes unnoticed by proponents of psychological mechanism.

Truth, Lies, and Self-Reference

What we have said so far about the paradoxical or self-invalidating nature of mechanistic explanations of the human psyche can be examined in general form, as a matter of logic.

In what might be called self-consistent statements, the circumstances of their utterance lie in the background. Thus, to say "The sky is blue" is to make a self-consistent statement. Nothing about its message runs counter to the fact of its being spoken. In this, it resembles most sentences and most beliefs, such as "It is raining," or "The road is black." All such self-consistent statements ride on a current of trust in the correctness of their content, the

sincerity of the speaker, and the capacity of the hearer to understand.

Now consider the sentence, "I am telling the truth." Unlike the simple declarative sentences in the previous paragraph, this assertion of one's own truth-telling, and any other statement referring to the circumstances of its communication, makes a self-referential point. It refers, not to states or events external to the speaker, but to its own speaking. (In another sense, all statements are self-referential, see Novalis, 1798; Kuehlewind, 1982). Yet it is like the other examples of self-consistency in that we assume that the speaker of "It is raining" believes what he says and means to say it; the same holds for "I am telling the truth." This latter sentence also rides on a current of trust in the correctness of its content, the sincerity of the speaker, and the capacity of the hearer to understand. These qualities inhere in every real act of communication.

Sentences such as, "I am now lying to you," (other examples would be "This sentence is in Hungarian," "I am not here," "No one is typing this") have a quite different structure. They are self-referential statements that run counter to the circumstances of their own utterance. They are what might be called pragmatic paradoxes (Levinson, 1983). They are not simply contradictions (as is, for

example, "It is raining and not raining"), but rather they run counter to the "pragmatic" situation in which they arise as statements. As we have just seen, it is not the self-referential quality that makes them paradoxical, since there exist an infinite number of self-referential statements that are self-consistent (further examples include "I am typing this now," "This sentence has five words," "I am here").

What is "wrong" with those self-referential statements that are also pragmatic paradoxes? They are attempts to run counter to the trust in communication referred to above. Since every act of communication depends on these elements of trust, explicit denial of the same elements is paradoxical -- for these elements lie concealed within the very project of denial itself.

Let us try to approach nearer to the problem inherent in pragmatic paradox as such. Bertrand Russell was familiar with this issue, and his "solution" is instructive. He held that all statements proceed at two levels or refer to two types of contents. One is the level or type of the statement itself, what I have been calling the circumstance of its utterance, and the other level or type is the content of the statement, that which it means explicitly or to which it refers. His theory of types was a kind of philosophical <u>fiat</u>, in which he simply ruled that propositions and their

predicates inhabit different logical orders, and therefore a statement cannot, for example, predicate either truth or falsehood of itself. As A.J. Ayer points out, however (Ayer, 1982), this is a somewhat arbitrary proscription. It is by no means clear why self-consistent, self-referential statements (e.g., "This is an English sentence") should be thrown out of the logically admissable along with what I am calling pragmatic paradoxes (e.g., "I realize thought does not exist"). Russell himself seems to have been dissatisfied with this aspect of his work (Brown, 1972).

Nevertheless, Russell's contribution was fruitful in that he clearly delineated two levels of meaning, that of the utterance and that of its explicit message. Instead of accepting his dictum that statements cannot make self-referential points validly, since the predicator is supposedly of a higher semantic order than its predicate, we could differentiate within self-referential statements and separate those in which there is a harmony between predicator and predicate from those in which there is an inconsistency between these two.

Russell's rejection of self-reference is not alone in failing to make this distinction, and in lumping all self-referential statements together. The influential writings of Douglas Hofstadter (Hofstadter, 1980; Hoftadter and

Dennett, 1981) make a related error on a grand scale. Yet it is the opposite error: that of admitting as meaningful and useful all self-referential statements, whether self-consistent or paradoxical.

The varieties of paradox have been analyzed by Copi (1973), van Heijenoort (1967) and others. Statements whose paradoxical nature derives purely from their violation of the rules of mathematical formulation (such as the so-called Russell paradox or Burali-Forti's paradox) are called "syntactic paradoxes." If the paradoxical statement's truth or falsity is dependent on its reference, however, we have what is called a "semantical paradox" (for example, "The statement at the end of this paragraph is false.")

As we can see, the key element in this last paradox (examination of the quoted sentence reveals that if it is true, it is false, and if it is false, it is true) relates to its reference to itself. "Pragmatic" paradoxes, as discussed in this thesis, can be thought of as a subset of semantical paradoxes: they are those semantic paradoxes whose problematic nature derives from the larger, extra-logical circumstances of their utterance. It is in this family that we locate the famous paradox of the Cretan Liar: "All Cretans are liars," says a Cretan.

If what the Cretan says is true, then, since it is uttered by a Cretan, it is a lie and necessarily false. But

if it is false, then sometimes Cretans speak the truth, and so it may be true .... but then it is a lie again and false! What makes the statement problematic is the extra-logical, pragmatic fact of its being spoken by a Cretan.

We can step off the merry-go-round of paradox, and still retain the statement "All Cretans are liars," only if we change the pragmatic circumstances of the utterance. Thus, if the same sentence were spoken by a non-Cretan, its problematic or paradoxical quality would fall away. It would no longer be self-referential, and it would no longer be paradoxical. Or again, the paradox would disappear if it were spoken by a Cretan who meant by it that all Cretans tend to lie, but who also meant that they occasionally speak the truth. On hearing such a statement from a Cretan, it is this latter assumption that would generally be made by a listener. We normally adjust the content of an utterance with its pragmatic setting so as to avoid logical conflict between the two (Levinson, 1983). The listener would assume that the Cretan was making use of a window of truthfulness to discuss his and his compatriates general untruthfulness. The paradox really comes into being only when it is a Cretan who makes the statement, and makes it or means it in its strong form: "All Cretans always lie."

What becomes of this strong form of the statement (i.e., "All Cretans always lie") when spoken by a Cretan

ignorant of his birth? For him, at any rate, it is not paradoxical. He feels himself to be in the position of any non-Cretan observing Cretan mendacity. For us, however, listeners who know that the speaker is a Cretan by birth -- is the statement problematic for us? There are two possibilities. Either we also know that the speaker doesn't know he is a Cretan -- in which case we know his statement is simply false, and not paradoxical, since here at any rate is a Cretan before us telling the truth to the best of his knowledge (though not to the best of ours). Since he spoke in good faith, he was not lying, though this good faith utterance makes his generalization about Cretans false.

In the other case, if we do not know that he is ignorant of his birth, we assume that he is a self-aware Cretan uttering the famous Cretan Liar paradox. We would be wrong in this, however. To summarize this last condition: the speaker would incorrectly believe himself to be telling a truth about Cretans; we would incorrectly believe him to be uttering a paradox. Only a third group, who knew of the Cretan's ignorance of his birth and our ignorance of his ignorance, could survey the whole situation accurately.

It should be clear from these examples that our understanding of certain self-referential statements depends on the pragmatics of utterance. It depends on who says

what, and what else they know. In discussions of the Cretan Liar and other paradoxes, little is generally said about the role of the interlocutor. How does one respond, after all, to the Cretan Liar? To ask this and related questions would be to delve not only into how listeners understand and reply to paradoxical utterances; it would also open up the issue of the Cretan Liar's next statement. In a dialogue, the paradox could not persist, without a bafflingly unself-conscious Cretan, or an illogically compliant listener. In other words, can we imagine a Cretan who knows he is one, utters the statement in its strong form, and then waits for our simple assent?

When mechanistic explanations of the human psyche are promoted, they put us in the position of listening to Cretan lying. That is, there is a logical equivalence to the propositions, "I am a Cretan and Cretans always lie," and "I am aware of the merely apparent nature of awareness," or "I understand that my understanding is caused by my brain." We can recognize in the Physicalists' Oath another equivalent paradox: "I vow to demonstrate the origins of vowing in physical/chemical laws." Baker isolates, in another context, the point of equivalence between the theoretical positions and the famous logical paradox: "it

is inconsistent to avail oneself of what one denies" (Baker, 1987). As she notes, the prohibition of such inconsistency is independent of any particular theory of thinking. The Cretan makes use of his interlocutor's pragmatic assumption that what the Cretan says is true, while what he says explicitly addresses and denies that assumption. In a similar fashion, to say, "I understand that my understanding is caused by my brain" is to assume (through "I understand") a freedom in cognizing that the sentence itself explicitly addresses and denies. "I understand X" always implies that X was not imposed on the subject through chemical, unconscious, or other mechanistic forces; if it had been, "I find X in consciousness," would be more accurate. Still, the "finding" -- as earlier the "understanding" -- is believed in naively, as an undetermined, cognitively capacious activity and not a mere result or reaction.

In general, to appreciate the nature of self-contradictory statements we must develop a sense for the hidden trust, mentioned above, that resides in every human assertion or statement. We trust, when speaking or writing, that our interlocutor understands, or could understand, what we are saying. We also trust in the correctness of what we mean (this is true of our lying as well -- in which case we simply say something other than what we mean). This trust,

the positive element concealed within Jonas's largely negative thesis, inheres in the making of all statements. It is their background or condition. We cannot consistently deny this trust (although we are free to do so if, like the self-conscious Liar, we are content with theoretical inconsistency). The import of a statement, Russell's "lower order," can either coincide with communicative trust, or else attempt to move against it. This criterion differentiates self-consistent statements from pragmatic paradoxes.

Pragmatic Paradox and Correctness

When Marvin Minsky writes that, "(Y)ou can build a mind from many little parts, each mindless by itself," (Minsky, 1985), he is involved in another pragmatic paradox, though this may not be immediately apparent. His assertion that a mind is built of mindless units in interaction with one another, all ultimately the product of chance and necessity, is another form of the computer-model. Like Freud's sometimes strict mental determinism, Minsky's (or Johnson-Laird's, or Edelman's) neuronal determinism prides itself on sober, dry factuality, and an ascetic immunity to the temptations of what Baker calls "common-sense conceptions of the mental."

Yet Minsky has written his book, wants its conclusions recognized, and vaunts the superiority of his view over previously held "folk psychological" explanations of mind. If he is correct, however, if we have only chance and necessity (of various flavors) to thank for the purely mechanical activity of our brains and minds, then he might just as well have published and believed in the opposite assertion. It too would have come about the way it had to, proceeding by natural laws and the complex workings of the brain.

What makes any understanding into understanding is a certain quality of self-supervision, about which we shall have more to say in subsequent sections of this chapter. It seems we do not really believe that our understandings are externally driven, caused, or determined, by some other, non-understanding process. Understanders believe inevitably that the current moment of understanding is its own final reason for being, that it is transparent to itself. If it were not so, if it were the product of something else (e.g. chance and necessity), then each understanding (what looked like understanding to the subject) would simply arrive in awareness (what looked like awareness to the subject) from a "darker," non-transparent source, with no guarantee that it was correct or even had anything to do with the matter at

hand. If my apparent understandings are forced upon me, by neurotransmitter activity, for example, then each new one, each latest understanding, might be irrelevant, incorrect, or pure nonsense as far as I could know. How would I be able to tell the difference from relevance, correctness and good sense? For if the psyche is a mechanism my attempt to tell the difference would itself be a necessary or chance event, not one arrived at by insight. For example, if someone asks me where rain comes from, I might answer, "From the clouds." Yet in a mechanistic universe, where thought proceeds from causes unsupervised by the thinking self, how could I be sure that this, my own answer, was in any way related to the question at hand?

As a stone falls by the simple workings of gravity, a thought, on Minsky's view (representative for a large portion of cognitive science), falls out from highly complex neural mechanisms. It only seems to be insight or correctness or freely understood. In fact -- says this position -- each thought is strictly predictable, or perhaps probabilistic: perhaps it is the non-periodic outcome of a nonlinear equation in a chaos-psychology (Woldeck, 1989). But how could a conclusion, a view, a theory such as these be known as correct (or incorrect) in such a world? This theory itself, like any thought or theory, would simply come

about as it comes about, like the stone falling. There is no understanding human being who could be relied on to grasp its import or even, for example, to compare its implications with experimental results: who could be sure the comparison was done correctly? Any verification, like any understanding, would be merely a mechanical process, which might go astray.

Actually, however, in a mechanistic universe there would be no "astray," as there would be no "incorrect." These words inevitably invoke by implication the supervision of conscious awareness. In a mechanistic world, such supervision is absent: there is no one there to choose, know, feel -- certainly no one to either achieve or appreciate a correct psychological or scientific conclusion. If we want to continue to assert correct views and mistakes overcome, our most rigorous, "scientific" model of the psyche will still have to include insight as a primary, irreducible reality.

Pragmatic Paradox and Belief

Minsky, like Hofstadter, "believes" in his own view. He does not change his mind about it from the inception of the book through to the end. He arranges his chapters and forms his paragraphs in order to make a point, to express

insight and conviction. He finds alternative arguments unconvincing; he wants to delineate reasons for his own. This entire structure assumes that there is such a thing as a view, a belief, a conviction, available to be held, even defended. Beyond the problem just discussed as to correctness, therefore, and at an even more basic level, the mechanistic position faces the problem of belief itself. Lynne Ruder Baker, in a critique of physicalism called, significantly, Saving Belief (Baker, 1987), examines this problem in depth, discussing it in the terms of current cognitive scientific debates.

As Baker points out, the physicalist views must, to be effective, eliminate "content" from the mental machine they want to contstruct. That is, they must piece together from purely physical processes everything we normally consider mental content. The relationship, for example, between one thought and the next, or one behavior and the next, must be explained exclusively in physical terms. At the operative level of description appropriate to such a theory, there can be no "reasons" or "beliefs" or other mental content that would account for the next moment's behavior: rather, an organism's actions follow from strictly physical states and laws of interaction. Minsky's own theory of the mental machine, then, to be consistently

applied to itself, would have to acknowledge that its own creation is not the result of reasoned insight based on experimental data, but rather a necessary and therefore unevaluable product of his own brain chemistry.

Thus, mental content and belief resist physicalist elimination. The proponents of elimination, for example, believe in its validity. Nor does their belief lack content: it is a belief in the applicability of their model. The more vociferously they argue for physicalism, the more evident becomes the contradiction. Why would Minsky be concerned, why even publish, if there is nothing, strictly speaking, to believe or say? When writing his book, when agreeing with one theorist as opposed to another, he tacitly expresses his hope that the very "mindlessness" he sees at the root of our actions does not, at any rate, determine his own.

Baker's critique is aimed specifically at physicalism. Yet the points she makes apply equally well to any mechanism, whether physicalist or non-physicalist. The inevitability of belief in the current content of consciousness pervades any theory, and is only paradoxical in the case of mechanism. To explain this belief in terms of stimulus-response patterns or drive states and their vicissitudes -- though neither of these positions depends immediately on physicalism -- involves the same paradox.

Pragmatic Paradox and Seeming

The cognitive philosophers (e.g. Dennett, Churchland) dismiss the preceding appeal to content on the grounds that of course they, like all of us, act as if mental content were real and causative in its own right. Yet this is all appearance, not reality. They, unlike the layman, at least recognize the falsity of such notions. Really these mental contents, beliefs, acts of will and so forth are the underlying neurological or at any rate natural and mechanically determined processes to which their theories refer. The mental contents only seem to be real. Churchland uses an analogy: just as we now know that the color red is a certain range of frequency of photon emission, we will one day be able to say more specifically that each apparent mental act is the corresponding neurochemical process.[2]

Baker asks the crucial question: how is it that there remains something so mentalistic as "seeming" in the

description that was intended to remove the taint of mentalism from these "mental phenomena"? The mental content dismissed as illusion pops up again in another form, namely as this very illusion. An intentional state (seeming) is still present in the final, most reduced equation of reality. The only way to avoid this conclusion would be to see this illusion, the "seeming," as itself unreal, an illusion, being actually, say, a pattern of neural firings. But if the illusion is not a real event, only an appearance, we once again have a mental content to explain away: the seeming of the illusion! It is clear that there is no end here: mental content will pop up at every explanatory level. We cannot escape it by ascribing its appearance to anything else, anything other than mental content, since as we have seen it is the very act of explaining it away that makes it appear again, one level down.

We can ask the same question as Baker with another emphasis if we focus on the illusory nature of the human subject implied in mechanistic theories. If mental contents are determined by natural processes of any kind -- chance and necessity operating at the level of the brain, the unconscious drives, stimulus-response mechanisms, etc. -- then the self-directing and self-aware human subject is called into question. On the mechanist view, when you or I

consciously direct our activities or assert our insight or feelings, we are imagining the existence of a self or subject whose presence is a kind of illusion. Actually, this apparent subject only seems to be real; it is a construct, built up from quite mindless, subjectless components. So runs the argument from physicalism. Our question must be, to whom does this illusion of a self or subject appear? Clearly we are in the same position here as before. Just as mental content could not be eliminated by explaining it as mere illusion, the subject of mental contents proves similarly resilient. If we clean out all the subjects at one explanatory level, unmasking them as illusory, then by that very act we create a new, unquestioned subject at the next level down.

As we saw in the Introduction, the error Descartes made in his Discourse on the Method was to slide away from his initial insight and assume that he had proved the existence of a particular kind of subject and a particular religious faith. We need not suppose, however (and as clinicians we are unlikely to do so), that the subject and the content that escape elimination are the existing self-concept or ego-structure of the individual. Rather, some subject is proved by this method to be ineluctably present, contentful, irreducible.

An Analogy: Goedel's Incompleteness Theorem

According to the definition of mechanism suggested in the Introduction, many views of psychology and medical science qualify as mechanistic. That is, many, perhaps most, theories presuppose either principally finite or probablistic, non-contentual programs for the production of human consciousness. The types of concepts they use are not based on conscious mental events, but borrowed from other sciences. Yet from La Mettrie through the Berlin physiologists and more current conceptions, the exact formulation of a mechanistic (meaning-less and will-less) description has proved elusive, for reasons the previous sections of this chapter have discussed. Despite the adoption, at every stage in its development, of Western society's latest technological achievement as a model for the human mind (Volpert, 1988), we are always left, in the unexamined interstices of every theory, with the presence of someone to whom the illusions of mental events appear, as we are left with the content and claim to self-aware correctness of the mechanistic views themselves. In psychoanalysis, this has been ably discussed as anthropomorphism or the "homunculus problem" (Schafer, 1976).

The quest for a description of the human in terms derived from the natural sciences is closely related to a similar quest within mathematics and logic. It is even linked historically, since the rise of psychology as an experimental science dates from 1879 (the year Wundt set up his laboratory in Leipzig), which was the same year that Gottlob Frege published his Elements of Logic. The struggle within logic, initiated in part by Frege, was the attempt to arrive at a complete, consistent, purely rule-driven derivation of true arithmetical theorems from basic logical postulates without recourse to either the unarticulable intuition of the theorems or to their practical application in order to prove their validity. This was the explicit undertaking of Russell and Whitehead's Principia Mathematica of 1913. They hoped to reduce the truths of arithmetic to a formal and severely logical system, so that given a minimum of initially unproved but logically certifiable axioms, all subsequent arithmetical truths could be derived exclusively by the application to these axioms of formal rules of substitution, transitivity, identity, etc.

There is a close analogy between this project and that of the cognitive scientists who would make of the human being a highly complex computer program, that of the behavioral theorists who see the human in terms of givens

and influences, and that of the neurologists who would locate brain function at the basis of mental activity. The mathematical project in question, called in one of its forms the Hilbert program after the German mathematician David Hilbert, was meant to proceed by "draining the expressions occurring within the system of all meaning: they are to be regarded simply as empty signs." (Nagel and Newman, 1958). What Hilbert was trying to show was that this procedure of formalization would lead to a proof of the absolute consistency of arithemtic, that is, to a proof that its axioms could never lead to contradictory theorems. It was to be hoped that a fully axiomatized, formalized arithmetic would be not only consistent but complete, that is, that all true statements within the arithmetic would be derivable from the axioms according to established rules. This project is continuous with the mechanist attempt to show how psychological and physical mechanisms account for the contentual, willed appearance of consciousness, while naming as its source certain relatively non-contentual, unwilled causes (Baker, 1987). Given the initial biochemical "programming," many neurologists and cognitive scientists for example seek to show how we could derive all apparently intentional mental content from the formal workings of biological or computerlike models (e.g., McClelland et al.,

1986). Just as the theorems of mathematics were to be derived without recourse to their intuitive content or practical application, the appearances of conscious mental content are to be derived without recourse to their intentional connection to previous and succeeding contents, but by reference to elements that are in principle inaccessible to awareness, such as physical states and events.

Yet the mathematical version of this project failed. Kurt Goedel's 1931 paper, "On Formally Undecidable Propositions of Principia Mathematica and Related Systems," showed that any axiomatic system powerful enough to generate the theorems of arithmetic will necessarily contain unprovable -- not just unproved -- true propositions. To put it another way, he showed that if mathematics is consistent, it is incomplete, thus setting a limit to the possible success of the Hilbert program. A key facet of Goedel's procedure was the discovery of a technique for including metamathematical statements (statements about arithemetic) within the very arithmetic that they describe. He used this technique to derive a formula within his logical system that declares its own undemonstrability, and he showed that this formula is necessarily true, but cannot be demonstrated within the formal system (Nagel and Newman, 1958; Berlinski, 1988; Penrose, 1989).

Commentators like Hofstadter have emphasized to a disproportionate degree the single facet of Goedel's work that involves his inclusion of metamathematical propositions about the formal system within the system itself. For Hofstadter and Dennett, this property suggests a number of conclusions not present in Goedel's work. It seems to them to suggest, for example, that all meta-systemic statements, such as our own theorizing about the nature of mind, are somehow reducible to intrasystemic statements, and thus ultimately obey the rules of a mechanistic system. Our statements about the brain (metasystemic) are ultimately reducible to brain processes (intrasystemic). Goedel's theorem would thus be an arrow in the quiver of mechanism.

Yet the point of Goedel's paper and subsequent writings was just the reverse. Goedel's work was important precisely because it showed that a particular formalized attempt thoroughly to mechanize a complex human creation (arithmetic) could not succeed, even if metasystemic propositions are formulated in the language of the system. Instead, such complex creations constrain us to refer to meaning, intention, and content of a true but unproveable kind. All complex logical systems are open, in the sense that their true propositions cannot be exhaustively proved. The thinking mathematician has access to truth-judgments that cannot be formalized.

Nothing in Goedel's theorem, then, suggests that there are limits to arithmetic truths themselves, nor to our capacity to know or create them. Hofstadter (1979), like certain computer-model advocates investigated by Turkle (1984), nevertheless expresses a concern that Goedel's theorem implies limits to human knowledge, to self-knowledge, or to human self-consistency. This concern makes sense only if we realize that it arises exclusively for those contemplators of Goedel who are already a priori committed to the mechanistic view of the human being. If we already imagine we are contentless, rule-driven systems, if we already imagine for example that our minds are computers, then, taking a hint from Goedel, we worry that we are incomplete -- that we cannot supervise the circumstances of our functioning. Yet Goedel's work was intended by Goedel, and has been understood by others (Lucas, 1961; Penrose, 1989), to suggest precisely the opposite, i.e. that cognition must be understood as a real, irreducible entity -- in fact, as the realest thing we have. Once again, it is the question of reflexivity: we too easily forget, when looking at the limitations of some human creation, performance, or function (in this case, formalized arithmetic), that it is no one else -- only humans themselves -- who notice such limitations, as Goedel noticed

a loop-hole in the logic of Principia Mathematica. This achievement of noticing error is always already the achievement of surmounting it.

Goedel was so far from believing in the physicalist point of view that he declared himself a proponent of mathematical realism. That is, he felt that mathematical categories have real existence, and are not mere abstractions from sense-perceptible observation:

> Classes and concepts may ... be conceived as real objects ...existing independently of our definitions and constructions. It seems to me that the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence.
> (Goedel, 1944)

Many discussions of Goedel's undecidability proof include warnings against drawing rash conclusions (e.g., Nagel and Newman, 1958; van Heijenoort, 1967). They caution us not to conclude, for example, that intuition rather than formal logic lies at the basis of mathematical reasoning. Nagel and Newman even feel constrained explicitly to align themselves with the physicalist project: "Nor do the inherent limitations of calculating machines imply that we cannot hope to explain living matter and human reason in physical and chemical terms." (The bristling double negative here suggests how near to hand is the opposite conclusion). Like Goedel, however, we can hardly find in

his location of unformalizable access to unlimited mathematical truths any support for such physicalism, or any demotion of conscious mental events from real to phenomenal status.

Goedel's theorem, and the whole field of formal logic, remain subordinate to the larger questions of psyche and soma. Formal undecidability can be used and abused in the service of various polemics, but it bears at most a metonymic relationship to the mind/body question. Formal logic and mathematics are narrowly local achievements in comparison with the psyche's already demonstrated range -- to say nothing of real or imagined horizons. It has often been noted that psychoanalysis has "physics envy," and we have discussed the tendency of every age to compare human consciousness with that age's most advanced technological product. All such analogies will eventually prove inadequate, if not simply tedious, since every fine human achievement suggests a wealth that beggars the particular case, however fine (Emerson, 1841). Psychology needs to cultivate a sense for the absurdity of limiting ourselves to the rules or formulae that we ourselves have discerned or invented.

As with analogy, so with authority. The physical or theoretical invention of the human requires a human not

bound by its workings to invent, appreciate, regulate, and improve on it. Similarly, no appeal to authority can be final, since it requires us, the humans who would be subordinate to the authority, to evaluate and accept it. This achievement could hardly demand less of us than shouldering for ourselves the burden we thus hoped to displace (Kuehlewind, 1982). In other words, the appeal to famous philosophers or psychologists whose work supports the conclusions of, for example, the present thesis would be as specious as the appeal to computer performance as a model for human cognitive capacity. Neither move can effectively discharge or disguise the instantaneous trust in our own insight as a real process. From the standpoint of empirical idealism, the touchstone of validity must be nothing other than the living encounter with the circumstances of our cognitive acts. The anti-mechanist theoretical conclusion outlined in this chapter is therefore already its empirical confirmation.

Chapter 3: The Implications of Reflexivity

Retrospect and Prospect

The principal points of the first two chapters can be easily summarized. This will help us to look at the implications, both theoretical and clinical, of the central thesis of psychological reflexivity and irreducibility.

In Chapter One, key moments in the history of mind/body problematization were reviewed, and an alternative both to materialist monism and Cartesian dualism was found in the Aristotelian concept of the epistemikon. For Aristotle, the "part" of the soul capable of self-examination is itself necessarily formless. Aristotle's concept of formlessness here seems to be of a different kind from his concept of Forms. When the psyche knows any Form, or when it knows itself (for it is a Form), it relies on what we might, non-metaphorically, call an intelligent readiness for new cognition. Aristotle calls this capacity an entity, the epistemikon. He defines it as a locus within

the soul, giving us a kind of pictorial representation of the irreducible and undetermined cognitive capacity, itself non-spatial, that we can logically deduce as belonging to the psyche.

Chapter Two focused on the pervasiveness of mechanistic explanations in psychology and the paradox at the heart of strong mechanistic frameworks of explanation. Such frameworks depend absolutely upon the contentual, free and insightful mental life that their theories tend to relegate to secondary or illusory status. The viewpoints of Jonas, Kuehlewind and Baker, with their understanding of cognitive "suicide," were contrasted with those of avowed psychological mechanists -- principally within behaviorism and cognitive psychology. The reductio ad absurdum that we applied to mechanistic explanation prompted insight into the inevitability, in theory and practice, of a certain kind of radical self-trust. By no means all kinds of self-referential statements are logically inadmissable or psychologically suspect; the problem lies only with those self-referential statements that move against this self-trust by denying explicitly the supervisory self-awareness they cannot help affirming implicitly. This is the lesson of reflexivity: it is we who comment on or theorize about our own psychological functioning, and we must therefore be

able to apply to every moment of commentary or theorizing whatever we claim applies to psychological functioning in general.

Chapter Two also clarified the issues of formalism and mechanism with regard specifically to Goedel's Undecidability Theorem. How Goedel's theorem came about, his own view of its implications, and its subsequent misunderstanding, are all suggestive analogies for the issues surrounding mechanistic explanations within psychology as a whole.

Chapters One and Two examined the ancient mind/body problem from a particular angle, and with a close focus. This perspective does not depend on seeing mind and body as separate substances in the manner of Cartesian dualism. Instead, it opposes a concept of <u>capacity</u> to all mechanisms, whether these latter are apparently mental or physical. Where proponents of all mechanisms see the human as essentially complete, definable, and caused (whatever probabilisms or randomizing functions they may include), the irreducibility perspective sees the human as having access to a radically undetermined, irreducible, intentional realm of formless capacity. It was to tease out the existence of this radical formlessness that we entered into the discussion of pragmatic paradox. Mechanistic models (for

example, the computer analogy or the neurological theory) fail to account for their own proclamation. They lay claim to contentual correctness while also undermining the contentual and the correct as such. Since, however, we cannot avoid our faith in correctness, content, and subjectivity, we have seen that we must acknowledge the presence in the psyche of an irreducible, non-mechanistic source of insight -- or renounce discourse altogether.

We take the result of the previous chapters as already accepted, for the purposes of exploring its implications and uses in the present chapter. This result can be stated in various ways. Most generally, we can say that understanding is fundamental. The science of psychology and the craft of psychotherapy depend on the use and investigation of the understanding, a full treatment of which can be complete without recourse to any non-cognitive elements. (Insight, cognition, self-trust, supervisory self-awareness, correctness, irreducibility, and other terms used so far have named the stages and conditions of what I subsume here under the general heading of the understanding.)

In the present chapter, we explore a few of the implications of this central thesis. A brief overview of the main points will follow, to prepare the way for their further elaboration later in the chapter.

The first and most important implication of the irreducibility thesis is simply to provide an apposite self-referential ground of explanation for existing techniques of psychotherapy. Whenever appeal is made to the patient's ability to choose, to understand, or to develop new capacities of any kind, the irreducibility thesis (or, in positive terms, the radical of self-trust and free capacity) is invoked by implication. For example, when we convey to the patient the importance of his or her commitment to treatment, to regular attendance, for example, we treat intentionality, purposiveness, and insight as real events. If attendance were exclusively a matter of unconscious drive states, stimulus-response mechanisms, or physiological processes, we would manipulate these processes rather than address the patient in words meant to be understood. The reliance on the free attentional capacity of understanding has been too obvious for remark, however, too blindingly simple, and has been overlooked in favor of mechanisms of all stamps as explanations for observed clinical phenomena and practice.

In this context, the similarity of mental "bodies" to physical bodies can be a useful analogy, giving the troubled term "mental energy" (Schafer, 1976; Mitchell, 1989) its

proper place as a heuristic image for attentional capacity. Diseased formations, like all existent patterns, belong to the mental body; the availability of free attentional capacity is increased when such formations are released in therapy. The psychic "energy" of attention, in turn, cannot derive from its own perceptions (e.g., from the physical body); if it has a source, the source must be even more cognitively active than itself.

A further implication we shall examine derives from simultaneously giving Ryle due credit for the strength of his arguments, analyzed in Chapter 1, on the "systematic elusiveness of the 'I'" (Ryle, 1949), while still holding to our irreducibility thesis. The conjunction of these two positions leads to a richer characterization of the epistemikon or unformed and undetermined aspect of the psyche.

Clinically, the implications of reflexivity for new treatments are twofold. On the one hand, we can understand the role of "insight" as the exercise of free attentional capacity, and we can extend our research and treatment strategies with such exercise in mind. On the other hand, the study of clinically well-known phenomena of psychosomatic disease and healing (c.f. McDougall, 1989) receives a new rationale. Perhaps the key theoretical

implication of the irreducibility thesis could be put as follows. Our actual patient, the real human subject, is not the person we see before us, neither the physical body nor the mental body as it is currently constituted; the real being of the patient is instead the patient's own capacity for emotional and cognitive growth.

Foundations of Psychotherapy

All psychotherapies make an appeal to the supervisory capacity of self-awareness that mechanistic theories disavow. The insight-oriented psychotherapies make this appeal explicitly (at least, intermittently) and locate the sources of self-supervisory awareness in both therapist and patient. The more strictly physicalist therapies (drug therapy, for example) locate the self-supervisory capacity in the therapist alone, but they cannot, as we have seen, eliminate it entirely. The most important implication of our irreducibility thesis is therefore to provide an understanding of the rationale behind insight-oriented therapies already in existence.

A paradigmatic instance of the appeal to supervisory capacity in psychotherapy occurred some years ago at the Harvard College Bureau of Study Council, where undergraduates go for counseling. The therapist asked the

patient, "Why are you here?" and received the answer, "My proctor sent me." To this the therapist replied with the interpretation, "So you decided to come."

This interpretive response, which therapists from many schools of psychology might make (on a good day), cannot consistently be uttered in a mechanistic framework. It points, in the style of the "reframing" interpretation, at the hidden purposefulness in what the student is doing. The student presents in effect a theory of motivation -- he claims to be like a billiard ball acted upon by another billiard ball, not a subject of his behavior but a helpless object, straight out of Galileo's Mechanics. Instead of accepting this mechanistic framework, the therapist offers an alternative theory of motivation that depends on actions freely chosen, alerting the patient to disavowed actions he has taken (c.f. Schafer, 1976). It also prepares a possible later interpretive focus on the patient's posture of mingled resistance to and compliance with authority, suggesting that this too is chosen.

The therapist's intervention here is not new in kind, but without an understanding of the reflexivity thesis it lacks theoretical support. In a mechanistic framework, the student's point of view would have to be acknowledged as closer to the truth than that of the therapist, or at least

as a more realistic kind of explanation for behavior. The therapist, in a mechanistic world, would be promulgating an instrumentally useful, but by his own lights false, sense of choice and insight. Having arrived at a reasoned understanding of the fundamental, irreducible nature of insight itself, we can make such interpretations with the conviction that they correspond to scientific actuality -- not merely to derivative, secondary, or instrumental processes. If aware of supervisory self-awareness as real, we would be more likely to develop techniques for fostering its growth.

Various methods have already been devised to prompt awareness of the understanding and choicefulness that the student in the above example had disavowed. D.W. Winnicott, in his paper on the False Self and the True Self (Winnicott, 1960), mentions how he sometimes had to stimulate a patient's awareness of his very existence by initially agreeing explicitly with the patient's existential doubts:

> The patient's False Self can collaborate indefinitely with the analyst in the analysis of defences....This unrewarding work is only cut short profitably when the analyst can point to and specify an absence of some essential feature: 'You have no mouth,' 'You have not started to exist yet' .... (T)he False Self, however well set up, lacks something, and that something is the essential central element of creative originality. (Winnicott, 1960)

The technique here, as in some family or systems therapies (Hoffman, 1981) is paradoxical in a therapeutic sense of the

term. It brings to the patient's awareness a psychic presence by ostensibly insisting on an absence. As the context of this interpretation in his article makes clear, Winnicott hopes for, and achieves, a counter-movement on the part of the patient in reaction to the interpretation.

Such interpretations are themselves miniature reductio's of the kind practiced theoretically in Chapter 2. They bring the patient's own mechanistic theory to a point of such absurdity as to evoke the presence of what was initially disavowed. Indeed, we can read the whole physicalist and mechanist tradition within psychology in the same way: as a grand, culture-wide reductio whose counter-movements have been various and always partial. Thus, for example, the cognitive revolution of the 1970's was the reaction to a stifling behaviorism, making the concept of mind respectable once again -- but at the cost of attributing mental events to the more fashionable mechanisms of the computer (Gardner, 1985). Perhaps as we collectively put the case of physicalism strongly enough, there will emerge, here and there, the healthy reaction of insight into irreducibility and psychic self-trust. When the old cannot die and the new cannot be born, however, a host of intermediate conditions will arise. Thus Winnicott's own case represents the variety of compromise

possible within this arena. He repeatedly stated his allegiance to a physiologically rooted psychic determinism, while his clinical writings suggest quite other conclusions, as in the technique quoted above (c.f. Winnicott, 1970; Mitchell and Greenberg, 1983).

Mental Bodies and Physical Bodies

A common characteristic of widely divergent mental and physical illnesses is their persistence. If dysfunction did not persist, no one would seek help for it, no doctors or psychologists would be called upon, and little would be written about it. Every disease is noteworthy for this simple quality of its persistence in the face of (someone's) dissatisfaction with it. It is precisely the quality of its not being amenable to conscious control that characterizes it as disease. The DSM-III-R definition of substance dependence disorders, for example, points up this quality when it defines substance dependence in terms of one's inability to discontinue substance use despite a desire to do so (American Psychiatric Association, 1987).

A classic psychodynamic characterization of common neuroses, D. Shapiro's Neurotic Styles (1965), shows in detail how in neurosis, as in substance abuse, it is the persistent styles of behavior, thought and feeling that

define disease. Whatever is seen to persist must have a specific and visible (noticeable) form. Rather than focusing on the causes, which seem various, Shapiro takes a stylistic or formalist view of neuroses and notes the various kinds of rigidity of the hysteric, compulsive, narcissistic and other patients. In a later book, Autonomy and Rigid Character (1981), Shapiro further describes how "autonomy," the capacity for variable focus and flexible response, is precisely the absent factor in, and antidote to, much mental illness.

Shapiro's sketches of neurotic style amount to descriptions of what I shall call diseased mental bodies, in recognition of their similarity to diseased physical bodies. Just as physical diseases possess a certain homeostatic persistence, drawing the body's energy to set patterns of dysfunction, so too we can think of a mental body, Aristotle's psyche, as a homeostatic mental organization where free energies can be mired in fixed, therefore diseased, mechanisms of thought and feeling. On this view, there are two kinds of bodies to be considered, mental and physical, and a noxious persistence, an entrapment of "energy", is the principle characteristic of both of their ailments.

This analogy bears a close relationship to our previous examination of reductive theses and their

deficiencies. We know that it is pragmatically contradictory to believe in a theory that makes of the mind a pure mechanism. From the standpoint of the energy analogy just mentioned, we would say that mechanistic psychologies attempt to cast the characteristics of disease over the whole of human consciousness. They correctly notice reflex-like, program-like, or chemically induced conscious states within human functioning, and then globalize these unhealthy states as if they alone accounted for mental life. This is why it was possible for Hofstadter (1979) to suppose Goedel's theorem suggested a limit to human creativity or self-knowing capacity. If we (but only if we) assume a priori that mechanistic laws arrange from the "inside" for what appear on the "outside" as conscious contents, the limits of formalism equal the limits of consciousness itself. If, however, the paradox of mechanism is recognized, we see instead that a relatively finished, fixed set of mental patterns constitutes merely a mental body, but not the psychic whole of us.

On this view, the ancient dichotomy of mind and body is thus replaced with a new distinction between relatively fixed, non-cognitive elements (both physical and psychological) on the one hand, and relatively free, cognitive elements on the other. Both have equal status as

objects of scientific study, though the means for examining the free cognitive capacities would have to include new techniques of introspection. Such techniques would look within cognitive capacity itself (in introspection therefore) for the sources of conscious mental life.

The relative freedom that makes one pole of this dichotomy is, at the same time, that "autonomy" recognized by Shapiro to form the only possible contrast to the fixed styles of mental illness. We have seen that this relative autonomy of the subject is not a matter of faith, but a question of insight into the circumstances of our speaking about. It is a truism of psychodynamic psychotherapy that the patient's most curative moment is the decision to enter treatment: the effort comes from that portion of the human being not caught in rigid cognitive and emotional responses. The patient has recognized that he has a problem, and with that recognition he reveals his partial freedom from the problem. This capacity for insight into the existence of a problem belongs to the agency defined by Aristotle as the epistemikon, which is in, but not of, the mental body: it alone has no fixed form, and can thus recognize the qualities of the psyche for what they are in any given case.

Mental Energy: A Question of Translation

The mental "energy" to which reference has been made in the previous sections of this chapter, the relatively free cognitive capacity, is non-physical and non-spatial. Another name for it might be simply "attention." When neurotic styles govern us, our attention has become trapped in relatively fixed forms (Wheelis, 1973; Epstein, 1987). To the extent that we can liberate our attention and make our behavior responsive rather than reflex-like, we have liberated a mental somewhat -- which has sometimes been referred to as mental energy. This usage corresponds to the heuristic sense of the term in the following translation of Freud's "Psychoanalysis" (Freud, 1922):

> It may be laid down that the aim of the treatment is to remove the patient's resistances and to pass his repressions under review and thus to bring about the most far-reaching unification and strengthening of his ego, to enable him to save mental energy which he is expending upon internal conflicts, to make the the best of him that his inherited capacities will allow and so to make him as efficient and as capable of enjoyment as is possible. The removal of the symptoms of his illness is not specifically arrived at, but is achieved, as it were, as a by-product if the analysis is properly carried through. (quoted in Menninger, 1958)

Freud's reliance on energic terms was borrowed from Breuer and ultimately from Helmholtz, one of the originators of the Physicalists' Oath (Laplanche and Pontalis, 1973). Yet from the very start, Freud (1915) used "bound" and "free" energies in ways that suited his clinical observations

rather than the physical science to which their usage made appeal and from which it borrowed scientific respectability (Sulloway, 1979). The tendency of recent psychoanalytic writing has been to avoid the "reification" implied in such energic terms, as if analogy were so dangerous a pursuit that it should be forbidden altogether (Schafer, 1976; Spence, 1982; Mitchell, 1988). This is above all a question of metaphor and its vicissitudes: when the topic of discussion is sufficiently global -- our own conscious functioning as a whole, for example -- then we can only compare it <u>downwards</u> to lesser, more limited spheres, and the terms we pick will inevitably fail to encompass the totality to which we refer. A possible solution is to embrace metaphors or specialized terms knowingly, so that their applicability is understood to be limited.

In Freud's definition of the goals of analysis, a number of points coincide with the implications of the irreducibility thesis. There is the focus on liberation of mental energies, which we can read as the process of transforming trapped patterns of attention into free attentional capacity. This comes about through exercise, especially exercise of the supervisory awareness: when resistances are removed and the patient can "pass his repressions in review," he is exercising the capacity to <u>see</u>

what he had formerly been. Freud seems to be equating this capacity with the ego, the "I." There is the further suggestion that when the ego is thus strengthened, mental energy has been saved. This energy can now be put to use, not in the fixed forms of conflict, repression and resistance, but mobilely, according to the will of the subject. Furthermore, symptoms are not directly attacked -- nor, indeed, were the repressions and conflicts directly attacked. Instead, this strengthening of the autonomous capacities of the ego deprives the fixed attentional forms of their life, their psychic sustenance, and they collapse or deflate on their own -- "as a by-product."

The useage of "mental energy" was meant by Freud, at least after the failure of his "Project" in 1896, to be a metaphorical or heuristic one (Grunbaum, 1984). Freud wrote,

> We assume, as other natural sciences have led us to expect, that in mental life some kind of energy is at work; but we have nothing to go upon which will enable us to come nearer to a knowledge of it by analogies with other forms of energy. (S.E.1940, 23:163-4 quoted in Gruenbaum, 1984)

Without endorsing the characterization of psychology or psychoanalysis as a "natural" science, we can applaud the restriction here of mental energy's field of reference to -- itself. No extra-mental analogies, Freud is aware at this point, will ultimately do. If there is to be a mental

energy, it borrows only the name, and no other quasi-physical elements, from other energies. His insight here perhaps prefigures Schafer's (1976) dissatisfaction with the excessive reifications of ego psychology.

In fact, "energy," "expense," and similar terms were often used quite interchangeably by Freud, so that English translators sometimes offer one for the other, as in the quotation above on the goals of psychoanalysis. For purposes of discussion I have left Strachey's translation intact; the original includes no reference to Energie, but only to psychische Aufwand ("psychic expense") (Freud, 1940). It is a question, more literally, of avoiding psychic expenditure rather than of saving psychic energy. Freud's use of an economic instead of a physical force metaphor, like Strachey's comfortable translation in terms of "energy," reflects the heuristic nature of the language used at the time.

Despite this sometime acceptance of the metaphorical limits of the energic model, Freud continued to oppose "bound" to "free" energies, with free energies characterizing the primary and bound energies the secondary process (Laplanche and Pontalis, 1973). Our own reasoning, however, as well as Freud's 1922 statement on psychoanalytic goals, suggests that a certain kind of free attentional

capacity (or energy) is also characteristic of what might be called tertiary process, or the post-analytic stage. What has been saved from neurotic defenses and conflicts serves the patient's variable and variegated healthy activities. It has become "free," either for the first time or once more.

If Freud initially, and then intermittently, linked the sources of mental life to the mechanistic biology of his day (Mitchell, 1988), he only participated in what we have discussed as the tendency of each age, continuing into cognitive psychology today, to pay for its mentalisms with selected mechanisms in the attempt to retain the claim to scientific standing. The sources of psychic energy or attention, are, however, non-physical -- a conclusion we derived from the logical inconsistency of calling them physical. We have more reason than Freud, then, to abandon the attempt to tie the fluctuations of conscious mental life to neurology. If we continue to use mental energy as a term at all, it is in full recognition that it bears few of the properties of physical energies, and is equivalent to or derived from none of them. We can read the references to "energy" in the psychoanalytic literature with more understanding, however, once we recognize that "attention" will often be an apt translation of the term.

Two possibilities remain. We can simply forsake the search for "sources" of the supervisory capacity of the psyche, or we can locate them in a still more understanding, more cognitively aware sphere than that of normal conscious awareness (Paul, 1972; Kuehlewind, 1982). This latter strategy would give us a tripartite model of the psyche, with unconscious, reflexlike tendencies "below," a conscious layer characterized by a mixture of autonomy and mechanism, and a supraconscious cognitive sphere from which normal cognitive processes derive (Kuehlewind, 1982).

Rylean Objections

It would be well, in this discussion of implications, to give due space to possible Rylean objections to the reflexivity thesis. In general, Ryle's point was that our location as subjects of our own discourse blinds us to the predictability, the pure natural scientific lawfulness, of our speech and actions. We can, Ryle notes, often see causes for other people's actions and for our own past actions. By extrapolation, then, we should acknowledge that even our own current awareness and thinking are proceeding according to essentially predictable laws.

In light of the discussion of mental bodies and mental energy, we can approach these objections in a different way

than was possible in Chapter 1. There, replaying the key move of our own reflexivity thesis, we noted that at the moment of making his own points, Ryle seems to hold to their correctness, cogency, and cognitive clarity in naive mentalist fashion. This observation repeated our commitment to questioning the circumstances of current utterance.[3]

Now, however, with a variegated model containing fixed attentional pattern on the one hand, and free attentional capacity on the other, we have a slightly different view of the matter. It appears that attentional movement itself, rather than the contents of consciousness, is the function that escapes mechanist formulation. We may indeed see our own past actions, as Ryle suggests, in terms of their incorporation within the mental body of relatively fixed contents. Nevertheless, the instant of moving from one content to its successor -- especially in the paradigmatic case of recognizing error -- continually regains our (if only momentary) allegiance. There thus emerges a hierarchy of relative freedoms and relative determinisms: while inevitably trusting the current moment of intentional awareness, we are free to recognize it as wrong the very next moment -- to see through it, to perceive its relatively limited, determined, erroneous nature.

There can be no final safeguard, no ultimate certainty, in this process, but neither can there be an abdication of certainty. For the very act of seeing the error or predictability of the previous mental moment reinstates self-trust. Our psychological insights must therefore be acknowledged both as free and motivated, both real and tentative. We are left with a psychological epistemology in continual flux -- unlikely to be welcome in a field that often seeks to impale ambiguity on the stake of natural scientific method. Instead of psychological theory and psychotherapeutic practice receiving some final formulation, we must accept the provisionality of our insights and become reconciled to our theory's perpetual revolution.[4]

Implications for Healing

In the last decade, popular (Cousins, 1979; Borysenko, 1988; Justice, 1989) and scholarly (Ader, 1981; Ader and Cohen, 1985) publications have created a new awareness of the potential for mind/body healing of many physical conditions, even of life-threatening physical ailments. Having seen the irreducibility of mental to physical events, and the common characteristics of mental and physical dysfunction, we can understand certain aspects of this phenomenon from a standpoint rarely taken in either the scholarly or popular literature. To do so, we need only take the full measure of the irreducibility thesis outlined in the preceding chapters. It will not, by itself, "solve" the psychophysical problem, nor give a full account of these healings. It will, however, begin an understanding analysis of mind/body healing, and suggest new treatment modes and avenues of research.

Some mental processes, we have suggested, do indeed operate quasi-mechanically. Habits of thought and feeling, repeated sensitivities and associations, may be adequately described by mechanistic models. These tendencies, like political "knee-jerk reactions," deserve comparison with physiological reflexes; they are predictable and

involuntary. Yet we have also seen that there not only are but must be other, voluntary and unpredictable, processes in the human being. These cannot be logically thought of as determined by any mechanistic substrate, whether physical or non-physical. But if these processes are independent of any mechanistic substrate, they are in particular free of the diseased mind or body of the individual human organism. They are therefore in a position to heal: they have radical independence and at least equal ontological status. If we search for what makes psychophysical healing possible, the unformed attentional capacity is thus a good candidate.

In fact, examination shows that it is these relatively free processes that are called upon in all psychophysical as well as psychological healing. From the existing literature, it is clear that in most cases of healing, a new capacity is inaugurated and exercised. Not a new habit alone, or a new activity, but new capacities are exercised. Exercise, in this sense, includes both repeated intentional practice and growth in expertise. Whether it is laughing (Cousins,1979) or visualization (Simonton, 1975) or the writing-out of painful memories, or simply the exercise of optimism (Justice, 1988), a field of activity new for the individual in question becomes a central focus of the therapy. A conscious effort is made to leave behind the

existing mental/physical body and develop attentional capacity (Kuehlewind, 1982). Healthy individuals operate along these lines without therapeutic intervention: for example, those people most resistant to the negative effects of stress (Kobasa, 1979; 1984) tend to be the ones who already possess a facility for greeting new situations flexibly, not grabbing onto existing securities or habits but open to creating new responses and engaging in new activities.

The implication from the techniques found effective within psychoneuroimmunology seems to be that our relative cognitive freedom has physiological consequences. This is only surprising in a mechanistic system, where cognition has not yet received, or reclaimed, the status of substance now widely accorded only to the physical body. Since the only evidence for physical bodies comes from our knowledge of them, however, (to what other source could we have access?) we can profitably question any system that grants scientific primacy to the products of cognition over cognition itself.

Psychoanalytic discussions of psychosomatic effects (McDougall, 1989) assume an interaction between unconscious meanings and physical ailments, but without giving an account of how this connection could occur. McDougall's presentation is suggestive, however, in that the

physiological consequence of unrecognized or unconscious psychic states tends to be noxious, while the physiological consequence of uncovering or reclaiming mental contents tends to be beneficial. As we discussed with regard to Shaprio's autonomy, insight liberates something -- we called it psychic energy. It now appears that what it liberates simultaneously relieves both the mental and physical bodies.

We have seen how hard it is to focus on a mental somewhat that is not a content, not formalizable in set rules, but that is rather the very condition of all our speaking, thinking, and rule-making. In the previous chapter, the existence of this capacity was highlighted by contrast; we saw how it is present even when denied, and perhaps most obviously present in those cases. Just because it is so difficult to focus on conceptually, it is hard to exercise and increase this mental somewhat. We can hardly exercise what we barely perceive. This accounts for some of the notorious elusiveness of psychophysical healing, its rarity and difficulty. Advances in this area will most likely come from the investigation of cognitive capacity itself, rather than of the contents of cognition (e.g., emotional conflict).

Rather than establish a kind of psychophysical pharmacopia, in which specific mental gymnastics would be

seen as curative of specific bodily conditions, the point here has been merely to draw out one possible line of therapeutic and theoretical implication from the thesis of psychological irreducibility.

Conclusion: Psychotherapy's Choice

The fruits of mechanistic psychology can be widely seen today. Sherry Turkle has written a sociology of the early computer culture in the United States (Turkle, 1984), and indicated the psychosocial effects of believing that people are machines. Her conclusions are even-handed. She notes how the growth and widespread availability of computer technology has alerted even toddlers to philosophical questions about the nature of mind that their parents perhaps asked only as adults. Yet she also notes the schizoid, depersonalized world of the computer-hacker and the tendency toward disavowal of personal and interpersonal responsibility among those who see themselves and others as computers. Turkle sees how the computer model performs in our intellectual life some of the functions earlier assumed by Freud's drive theory. It questions the sovereignty of the conscious "I" and offers a disarmingly deterministic model in place of common sense conceptions of mentality.

Dreyfus and Dreyfus (1986) have extended these observations by pointing out how the model of human

cognition and performance emerging from the artificial intelligence community has influenced other portions of society, causing an over-reliance on the potential of machine-helpers and an abdication of human cognitive potential. Calculative rationality, as they put it, becomes the self-limiting, self-predictive model for all human creative ability. In medicine, there has been a widespread deterioration in doctor-patient relations, which derives at least in part from the physicalist schemas of doctors who have been trained to see their patients as malfunctioning machines (Calahan, 1990).

The mechanist model does seem to have identifiable and harmful social consequences. It is harder to point out the consequences of mechanistic as opposed to non-mechanistic models specifically for psychotherapy because of inconsistencies in the application of theory. We have determined that a consistent mechanist would have very little to say to the patient -- very little to say at all (Baker, 1987). Yet as soon as psychotherapy begins, a world of belief, insight, cognitive freedom and form-free capacity is implicitly invoked by the therapist, of whatever doctrinal orientation.

When the patient's form-free capacity is exercised, the process has healthful consequences both psychically and

When the patient's form-free capacity is exercised, the process has healthful consequences both psychically and physiologically. The endeavor of the present thesis has been to show the scientifically fundamental status of mental content and to suggest that it is the exercise of unformed cognitive capacity that accounts for what psychotherapies already demonstrate in the way of success. Rather than experiments that look outward at observable third-party events, one important line of psychological research will involve new introspective methods to investigate the sources of cognition within, rather than outside, cognition itself.[2] The resulting development of new psychotherapeutic techniques for the exercise of cognitive capacity would improve psychotherapy to a degree, and in directions, that are of course impossible to predict.

1An interesting variant on this abdication emerges occasionally in physicalist, often cognitivist, schemas. it is the attempt to forestall the task of insight by asserting a Reymondian ignorabimus in contemporary form: "Consciousness may be determined, it may be undetermined; we cannot know which view is correct" or, "We await proof of the reality of mental events." As Johnson-Laird puts it (Johnson-Laird, 1988), "Nobody knows what consciousness is for...." All these statements of course involve profound commitment to the reality, correctness and freedom of mental content.

2Churchland's example (Churchland, 1984) begs the question, since the point at issue in color perception is the same as for other cognitive processes. From the standpoint of a psychology of perception consonant with the implications of the present thesis, color does not reduce to electromagnetic frequency.

(continued)

3The fact that others elsewhere or at another time may detect mechanistic causes behind our current utterance does not change the situation fundamentally, for each of these others' assertions is also intended, believed in, and considered with non-mechanist conviction at the moment of insight or utterance. At each remove to an imagined level from which the examined activity appears lawful or predictable, a law-detector or predictor appears whose own activity goes unquestioned.

4This is the process whose appeal, but not whose point, is reflected in the self-conscious gesture of Lawson (1985) and others who almost, but never fully, apply their doubts about referential language to their own moments of reference. The doubt or suspicion (which is also at the source of all mechanisms) comes from a dissatisfaction (justified from another standpoint) with naive realism and naive linguistic reference. It properly hovers over all previous insights and statements, but as we have seen cannot attack the current instant of insight or the current statement.

5Baker's (1987) insight that in a physicalist system the connection between thoughts would have to be explained in purely physical terms gives a clue as to how such research would proceed. It would not seek a mechanist source, but find within thinking itself the initially unobservable supporting medium between successive thoughts. Such a procedure, of concentrated inner attention, is prefigured in Marion Milner's perceptual experiences (1981), in Kuehlewind's suggestions (1982) and in some of the "flow" research of Cziksentmihali (1990).

Bibliography

Ackrill, J. (ed.). (1987). A New Aristotle Reader. Princeton, NJ: Princeton University Press.

American Psychiatric Association. (1987). Diagnostic and Statistical Manual of Mental Disorders. (Third Edition - Revised. Washington, DC: American Psychiatric Association.

Atkinson, R., Atkinson, R., and Hilgard, R. Introduction to Psychology. New York: Harcourt Brace Jovanovich.

Ader, R. (ed.) (1981). Psychoneuroimmunology. New York: Academic Press, Inc.

Ader, R. and Cohen, N. (1985). "CNS-Immune system Interactions: Conditioning Phenomena". The Behavioral and Brain Sciences. 8,379-94.

Ayer, A.J. (1982). Philosophy in the Twentieth Century. New York: Random House.

Baker, L.R. (1987). Saving Belief. Princeton: Princeton University Press.

Barrett, W. (1987). Death of the Soul. New York: Anchor Books.

Berkeley, G. (1965). Berkeley's Philosophical Writitngs. New York: Collier Books.

Berlinski, D. (1988). Black Mischief: Language, Life, Logic, Luck. New York: Harcourt, Brace, Jovanovich.

Born, R., ed. (1987). Artificial Intelligence. London: Croom Helm.

Borysenko, M. (1987). "Area review: Psychoneuroimmunology". Annals of Behavioral Medicine. 9(2) 3-10.

Borysenko, J. (1987). Minding the Body, Mending the Mind. New York: Bantam.

Brady, R. (1981). "Mind, models and Cartesian observers: a note on conceptual problems". Journal of Social Biological Structure. 4,277-86

Brown, G. (1972). The Laws of Form. New York: The Julian Press.

Calahan, D. (1990). What Kind of Life?. New York: Simon and Shuster.

Churchland, Paul. (1988). Matter and Consciousness. Cambridge, MA: MIT Press.

Copi, I. (1973). Symbolic Logic. New York: Macmillan Publishing Co.

Cousins, N. (1979). Anatomy of An Illness. New York: W.W. Norton.

Descartes, R. (1975). Discourse on Method. F.E.Sutcliffe, trans. London: Penguin.

Discover (May, 1989). "Pavlov's Rats." in "Up Front".

Dreyfus, H. and Dreyfus, R. (1986). Mind Over Matter. New York: The Free Press.

Emerson, R.W. (1841). "The Over-Soul". Essays, First Series. in Ralph Waldo Emerson: Essays & Lectures. Porte, J., ed. New York: Library Of America.

Flanagan, O. (1985). Consciousness, Naturalism and Nagel. Journal of Mind and Behavior. (6) 373-90.

Frank, J. (1973). Persuasion and Healing. Baltimore: Johns Hopkins University Press.

Freud, S. (1953). The Interpretation of Dreams. Strachey, J., trans. London: Hogarth.

Freud, S. (1940). Gesammelte Werke. London: Imago Publishing Company Ltd.

Fromm, E. (1941). Escape From Freedom. New York, Rinehart and Company.

Gardner, H. (1985). The Mind's New Science. New York: Basic Books.

Gazzaniga, M. (1988). Mindmatters. Boston: Houghton Mifflin.

Gleick, J. (1987). Chaos. New York: Penguin Books.

Goedel, K. (1980). Collected Works. Vol. 1 (Publications 1929-1933 New York: Oxford University Press.

Greenberg, J. and Mitchell, S. (1983). Object Relations in Psychoanalytic Theory. Cambridge, MA: Harvard University Press.

Gruenbaum, A. (1984). The Foundations of Psychoanalysis: A Philosophical Critique. Los Angeles: University of California Press.

Hampton-Turner, C. (1981). Maps of the Mind. New York: Collier Books.

Harre, R. (1985). The Philosophy of Science. Oxford: Oxford University Press.

Haugeland, J., ed. (1987). Mind Design. Cambridge: MIT Press.

Heilman, K. and Valenstein, E. (1985). Clinical Neuropsychology. Oxford: Oxford University Press.

Hoffman, L. (1981). Foundations of Family Therapy. New York: Basic Books.

Hofstadter, D. & Dennett, D. (1987). The Mind's I. New York: Bantam Books.

Hofstadter, D. (1980). Goedel, Escher, Bach. New York: Vintage Books.

Irani, K.D. (1977). "Conceptual Structures Underlying Types of Explanations in Psychopathology". Science and Psychotherapy. Edely, et al., eds. New York: Haven Publishing Company.

Irani, K.D. (1978). "Explanations in Psychoanalytic Theory". Pathology and Consciousness. Irani et al., eds. New York: Haven Publishing Co.

Irani, K.D. (1980). "Conceptual Changes in the Problem of the Mind-Body Relation" in Rieb, R. (ed.) Body and Mind: Past, Present and Future. New York: Academic Press, Inc.

Jennings, J. (1985). The Fallacious Origins of the Mind-Body Problem. Journal of Mind and Behavior. (6) 357-372.

Johnson-Laird, P. (1988). The Computer and the Mind. Cambridge: Harvard University Press.

Jonas, H. (1987). Macht oder Ohnmacht der Subjektivitat? Frankfurt: Suhrkamp.

Justice, B. (1988). Who Gets Sick: How Beliefs, Moods, and Thoughts Affect Your Health. New York: St. Martin's Press.

Kobasa, S. (1979). "Stressful life events, personality, and health: An inquiry into hardiness. Jopurnal of Personality and Social Psychology. 37(1),1-11.

Kobasa, S. (1984, September). "Test for Hardiness: How much Stress can you Survive?". American health, p.64.

Kolb, B. and Whishaw, I. (1985). Fundamentals of Human Neuropsychology. New York: W.H. Freeman.

Kuehlewind, G. (1982). Das Leben Der Seele Zwischen Ueberbewusstsein und Unterbewusstsein. Stuttgart: Verlag Freies Geistesleben.

Lange, J. (1970). The Cognitivity Paradox. Princeton: Princeton University Press.

Laplanche, J. and Pontalis, J-B. (1973). The Language of Psychoanalysis. New York: W. W. Norton.

Levinson, S. (1984). Pragmatics. Cambridge: Cambridge University Press.

Lucas, J. (1961). "Minds, Machines and Goedel". Philosophy 36 (1961).

Luria, A.R. (1966). Higher Cortical Functions in Man. New York: Basic Books.

McDougall, J. (1989). Theaters of the Body. New York: W.W. Norton.

Menninger, K. (1958). Theory of Psychoanalytic Technique. New York: Harper and Row.

Milner, M. (Joanna Field). (1981). A Life of One's Own. New York: J.P. Tarcher. (originally published 1955).

Minsky, M. (1985). The Society of Mind. New York: Simon and Shuster.

Mitchell, S. (1988). Relationsl Concepts in Psychoanalysis. Cambridge, MA: Harvard University Press.

Nagel, E. and Newman, J. (1958). Goedel's Proof. New York: New York University Press.

Nagel, T. (1974). What is it like to be a bat? The Philosophical Review.

Novalis. (1798). "Monolog". in Novalis. (1975). Reinbek bei Hamburg: Rowohlt.

Ostenfeld, E. (1986). Ancient Greek Psychology. Denmark: Aarhus University Press.

Pagels, H. (1988). The Dreams of Reason. New York: Simon and Shuster.

Pagels, H. (2-19-88). The Computer Scapegoat. New York Times. Op-Ed.

Paul, I.H. (1972). Letters to Simon: On the Conduct of Psychotherapy. Madison, CT: International Universities Press.

Peele, S. (1985). The Meaning of Addiction: Compulsive Experience and its Interpretations. New York: Lexington Books, 1990.

Pelletier, K. (1977). Mind as Healer, Mind as Slayer. New York: Delta.

Pelletier, K. & Herzing, D. Psychoneuroimmunology: Toward a Mindbody Model. Advances., (5) 27-56.

Popper, K. & Eccles, J. (1977). The Self and its Brain. London: Routledge & Kegan Paul.

Putnam, H. (1975). Mind, Language & Reality. Cambridge, England: Cambridge University Press.

Robinson, D. (1986). An Intellectual History of Psychology. Madison, Wisconsin: University of Wisconsin Press.

Rumelhart, McClelland and the PDP Research Group. (1986). Parallel Distributed Processing. vols. I & II. Cambridge, MA: MIT Press.

Ryle, G. (1949). The Concept of Mind. New York: Barnes and Noble.

Schafer, R. (1976). A New Language for Psychoanalysis. New Haven: Yale University Press.

Schorr, A. (1984). Die Verhaltenstherapie. Basel: Beltz Verlag.

Searle, J. (1984). Minds, Brains and Science. Cambridge, MA: Harvard University Press.

Shapiro, D. (1965). Neurotic Styles. New York: Basic Books.

Shapiro, D. (1981). Autonomy and Rigid Character. New York: Basic Books.

Simonton, O. and Simonton, S. (1975). "Belief Systems and Management of the Emotional Aspects of Malignancy". Journal of Transpersonal Psychology, 7,No.1,pp.29-48.

Simpson, D. (1989). "Going on about the War without Mentioning the War: The Other Histories of the 'Paul de Man' Affair". The Yale Journal of Criticism. Vol.3, No.1.

Singer, E. (1965). Key Concepts in Psychotherapy. New York: Basic Books.

Sperry, R. (1986). Discussion: Macro- versus Micro-Determinism. Philosophy of Science. (53) 265-70.

Spence, D. (1982). Narrative Truth and Historical Truth. New York: W.W. Norton.

Sperry, R. (1987). Structure and Significance of the Consciousness Revolution. Journal of Mind and Behavior. (8) 37-66..

Sulloway, F. (1979). Freud, Biologist of the Mind. New York: Basic Books.

Turing, A. (1950). "Computing Machinery and Intelligence". Mind, Vol. LIX, No.236.

Turkle, S. (1984). The Second Self. New York: Simon and Shuster.

van Heijenoort, J. (1967). From Frege to Goedel: A Sourcebook in Mathematical Logic. Cambridge, MA: Harvard University Press.

Vesey, G., ed. (1964). Body and Mind. London: Allen and Unwin.

Volpert, W. (1988). "Der technik-zentrierte Mensch und seine begradigte Welt". Mensch, Computer, Erziehung. Gergely, E. and Golmann, H. eds. Vienna: Boehlau.

Wachtel, P. (1977). Psychoanalysis and Behavior Therapy. New York: Basic Books.

Wheelis, A. (1973.) How People Change. New York: Harper and Row.

Whorf, B.L. (1956). Language, Thought and Reality. Cambridge, MA: MIT Press.

Winnicott, D.W. (1965). The Maturational Processes and the Facilitating Environment. New York: International Universities Press.

Winnicott, D.W. (1971). Playing And Reality. London: Tavistock Publications.

Wittgenstein, L. (1958). Philosophical Investigations. (Anscombe, G. trans.)
New York: Macmillan Publishing Co.

Woldeck, Rudolf v. (1989). "Formeln fuer das Tohuwabohu". in Kursbuch: Chaos. Nr.98, Nov., 1989.

Wortman, C. and Loftus, E. (1988). Psychology. New York: Alfred Knopf.

Yalom, I. (1980). Existential Psychotherapy. New York: Basic Books.