CLARK GLYMOUR

3 Freud's androids

A recent essay in *Science* compares Freud's work with contemporary "cognitive science." The comparison is rather to Freud's disadvantage, and to the disadvantage of Freud's contemporaries: *Our* contemporaries have a conception of the mind as a computational system. Some of their theories posit a quantity, "activation," that is responsible for aspects of mental functioning. Some of their theories postulate "parallel processing" through a network that is analogous to the connected system of nerve cells in the human nervous system. Unlike Freud, the story goes, our contemporaries have an experimental tradition that supports their theories. The result is that we now have a powerful and distinctive science of both the unconscious and the conscious, a science whose theories have led to new experiments "that tentatively reveal a tripartite classification of nonconscious mental life that is quite different from the seething unconscious of Freud."[1]

In a general way, these perceptions are widely shared, not only among academic psychologists, but among philosophers of mind, philosophers of science, research administrators, and increasingly, the educated public. They have the impression that contemporary cognitive psychology with its computer simulations of mind is onto something new and scientific that was at best only dimly foreshadowed in earlier psychologies. My purpose is to argue the contrary. A big part of contemporary cognitive science is pretty much what you would expect to get if Sigmund Freud had had a computer.

I thank Jerome Neu for helpful comments on this essay.

I

While the popularity of cognitive science, the digital computer, and the formal theory of computation are all relatively new, most of the basic ideas of contemporary cognitive science are *not* new. They appeared nearly in their present form in the late nineteenth century in the work of a group of neuropsychologists and neurophysiologists: Hermann Helmholtz, Theodor Meynert, Ernst Brücke, Jean-Martin Charcot, Pierre Janet, Carl Wernicke, Sigmund Exner, Joseph Breuer, and others. One of the others was Sigmund Freud. As an intellectual community, they were at once unified enough in theme and different enough in details to represent almost every fundamental idea of our own contemporary cognitive psychology. Freud and his contemporaries lacked the notion of a digital computer, of course, and of computation theory, and they also lacked the specific algorithms that have been proposed in the last thirty years to explain specific cognitive capacities. But they did not lack the idea that the brain is a biological machine that executes algorithms, nor were they without ideas about the computational architecture of that machine, nor did they lack the several conceptions of psychological explanation that are at work in contemporary sciences of the mind. Freud, especially, did not lack any of these things.

The neuropsychology of the late nineteenth century does not just anticipate our own; on the major conceptual issues it is quite as developed. Freud and his contemporaries understood the value of tying psychology to physics and biology, and they disputed among themselves the value of locating the mechanisms of thought in particular regions of the brain. Freud and his contemporaries understood the brain as a computational device, and they hypothesized a "language of thought" analogous to what we would nowadays call a "machine language" for a computer. They understood the elements of what we now call "connectionist" computation, and they made proposals as to how, using thermodynamic principles, connectionist devices can learn. Freud himself introduced much of the equivocal character that besets contemporary accounts of mental states as *functional* states. He employed a conception of homuncular explanation that anticipated contemporary modes of explanation in economics and political science, and that is philosophically unexceptionable. Freud's understanding of mental representation derived as much from the arts as from biology, and the arts provided him with a view about representation and rationality that has implications for contemporary discussions of the relation between rationality and analog computation.

Freud tended to exaggerate every intellectual issue, and, especially in his more youthful work, tended to look for unequivocal, radically general, and uncompromising formulations of fundamental hypotheses about the mind. A certain extremism is one mark of a philosophical intellect, for it tends to make issues stark and simple and as general as possible, the way philosophers like them. The result is that Freud's writings contain a philosophy of mind, and indeed a philosophy of mind that addresses many of the issues about the mental that nowadays concern philosophers and ought to concern psychologists. Freud's thinking about the issues in philosophy of mind is often better than much of what goes on in contemporary philosophy, and it is sometimes as good as the best. Some of it is dated, of course, by the limits of Freud's scientific knowledge, but even when Freud had the wrong answer to a question, or refused to give an answer, he knew what the question was and what was at stake in it. And when he was deeply wrong, it was often for reasons that still make parts of cognitive psychology wrong.

These claims may seem mysterious. Why, if Freud was a spokesman for a movement that almost fully anticipated contemporary cognitive psychology, is that fact not already recognized? And how did Freud come to be seen as the source of a movement, psychoanalysis, pretty much orthogonal to contemporary cognitive psychology? Cognitive psychology is a new discipline even if it is not a new subject. The parts of Freud's work that most clearly develop and illustrate the foundational issues in cognitive psychology were written before the turn of the century; they are unread by most academic psychologists, and they do not include any of Freud's most popular writings. It was in his early years, while still directly under the sway of the neuropsychological and neurophysiological communities, that Freud formulated the basic themes with which we shall be concerned. Psychologists, like almost everyone else, know Freud principally from a later period of his life; without the contrast of the earlier period of Freud's work, the issues that concern us are less vivid and more difficult to discern.

Sigmund Freud entered medical school at the University of Vienna in 1873. His medical education, which continued for eight years, was divided by two attachments. One, to Franz Brentano, the defrocked Catholic priest who had come to the University of Vienna as Professor of Philosophy the semester after Freud had begun his studies, occupied the first two years of Freud's career as a medical student. The other, to Ernst Brücke, Professor of Physiology, continued for the rest of Freud's student days and for some while after. Two other men, Theodor Meynert, Professor of Psychiatry at the University, and Josef Breuer, one of the most eminent Viennese physicians, also had powerful influences on Freud in his student years. Brentano on the one side, and Brücke, Meynert, and Breuer on the other, framed the understanding of mind and matter that Freud endorsed. The views of the two sides were very different in some important respects, alike in others, and where they differ Freud's opinion came to rest with Brücke's side rather than Brentano's.

Brentano gave Freud all the formal philosophical tutoring he was ever to have. Freud learned logic – Aristotle's theory of the syllogism – from Brentano, and he learned the strategems of philosophical argument. In 1874, while Freud was studying with him, Brentano published *Psychology from an Empirical Standpoint*, and the contents of that book gave Freud one vision of what psychology should seek to know, and of what methods it should use. Brentano's views of the goals of psychology were simple and rather traditional. Everyone has private access to one's own mental phenomena, to thoughts and dreams and images and pains and pleasures. To deliberately recollect one's own mental phenomena is to *introspect*. By introspection, properly conducted, everyone can collect facts about one's own mental life. The facts revealed to different people will of course be different, but according to Brentano there must be regularities revealed in any one person's mental life, and the regularities will be the same from person to person. Those regularities are the laws of mental life, and to find them is the proper goal of empirical psychology.

Brücke, along with Emil Du Bois Reymond and Hermann Helmholtz, had studied physiology with Johannes Müller. Müller was a sort of vitalist, who held that the workings of the body could not be entirely explained on physical and chemical principles. He must have wanted either in charm or persuasiveness, for history has it that his three most distinguished students allied themselves against his doctrines. Their views were essentially those of the great French physiologist Claude Bernard, who in 1865 popularized scientific materialism in his *Introduction to the Study of Experimental Medicine.*

The essential doctrine shared by Brücke, Du Bois Reymond, Helmholtz, and Bernard is what philosophers nowadays call the doctrine of *supervenience.* The idea is that one set of properties determines another set in every possible circumstance. Property P supervenes on a set S of other properties provided every pair of possible circumstances that are alike with regard to S are also alike with regard to P. The physiologists held that all properties supervene on the physical properties; same physics, same everything else. They also held to a strict physical determinism, by which they meant that if two systems should be in the same physical circumstances at corresponding moments, then those systems would also be alike in their physical states at subsequent, corresponding moments. The doctrines of physical determinism and supervenience evidently together imply that determinism holds for all properties of all things, not just for physical properties. Determinism and supervenience together promoted a contempt for statistical methods in science.

Brücke, Freud's most influential teacher, was a physiologist, and so were Du Bois Reymond and Helmholtz, his compatriots in the nation of materialism. Freud did anatomy with Brücke, chiefly neural anatomy, which was also one of Meynert's specialties. In Brücke's laboratory physiology and anatomy were one subject pursued by different methods. Physiology, like any other science, is many things. Traditionally it is the study of *functional structure* in living organisms. Theories of functional structure are really special kinds of *decompositions* of capacities. Humans live; how do they do it? They do it by eating and breathing and excreting. And how do they breathe? They do it by inspiring air into the lungs, absorbing part of it into the blood through the lungs, and expiring the remainder of the air and gases received from the blood. And how are these things done?

Physiological explanations do several things at once. They focus on a capacity to be explained, they decompose it into component capacities that together are supposed either to *constitute* the capacity to be explained, or to have it as an *effect.* But the component capacities are produced by specific *physical structures* within the organism. Breathing involves the nose and mouth, the larynx, the lungs, the diaphragm. In physiology, the analysis of functional structure is concomitant with the analysis and description of physical components that carry out the component functions or capacities. The connection of function and physical structure permits the order of questions to be reversed. When a new, discrete anatomical structure is discovered, one can ask what its function is, which is only a way of asking what capacities are based on the anatomical part.

Now the materialist school of physiologists held that the analysis of capacities ought to end in physics and chemistry. The capacity to breathe is analyzed into the capacity of the lungs to inspire, expire, and to exchange gases with the blood. The capacity to exchange gases with the blood is analyzed into changing physical conditions, namely the volume, pressure, and chemical composition of the gases in the lungs, the concentrations of various chemicals in the blood, the mechanical effects of increased air pressure in the alveolae, and the laws of thermodynamics and diffusion. In the end, nothing remains in any instance but physics and chemistry.

Materialist physiology, the sort of physiology advocated by Brücke and the other members of the Helmholtz circle, must inevitably be extended to a materialist psychology as well. The analysis of biological capacities must at many points appeal to capacities of the brain, and to *cognitive capacities.* Processes that appear to be under "voluntary" control must, according to Brücke and his colleagues, be analyzable into capacities that are finally explicable in physical and chemical terms. The cognitive capacities include the ability to recognize things, to locate them in space and to manipulate them, the ability to remember, to learn and solve problems, and above all, the ability to converse and communicate. Language seems a crucial case. If the capacity to communicate in language could be analyzed into component capacities, and ultimately into physical and chemical structures and processes, one of the great challenges to materialist physiology would be met.

How even to begin to construct a cognitive physiology? In ordinary physiology there are specific tissues involved, and one can use essentially physical experiments to examine the causal properties of those structures in order to discover the component capacities. But with cognitive capacities there is only one structure, the nervous system, and it is difficult to get at and to manipulate. Without such manipulation, it would appear that one can only guess at the component capacities that make up the capacity to converse.

Traditional philosophical psychology analyzed the mind into a collection of "faculties," the Will, the Imagination, Reason, Judgment, and so on. The faculties form a kind of organizational chart of the mind, with each faculty given a set of powers or functions. Faculty psychology is like physiology *without* physics. Two of the most powerful ideas in the theory of mind developed in the nineteenth century are that the traditional faculties are the wrong way to decompose human capacities, and that the right ways, the correct subcapacities, are based on specific tissues within the brain and nervous system. Francis Gall advocated the localization of faculties in regions within the skull, but the real advance in the idea of localization turned on novel analyses of the capacity for language.

In 1861, Broca claimed to have located a region of the cortex responsible for the *production* of speech. Stimulated by Broca's work, Theodor Meynert and his student, Carl Wernicke, began a kind of physiology of the mind whose signal triumph was announced in 1874, the same year in which Brentano's book was published, and the second year of Freud's medical studies.

Wernicke's triumph was the discovery of a region responsible for the *comprehension* of speech. The work was a combination of neuroanatomy and clinical psychiatry. Patients with linguistic *incapacities*, aphasias, were classified by the particular sort of incapacity they exhibited, and when the patients died their brains were examined for lesions. The location of the lesion identified the region of the cortex responsible for the patient's aphasia, and hence a region necessary for the corresponding linguistic subcapacity.

Meynert and Wernicke decomposed the capacity for speech into a set of subcapacities: the capacity to hear, the capacity to interpret sounds as speech and understand the speech, the capacity to reason and think, the capacity to produce speech. They supposed each of these capacities to have a physical locale in the brain; special tissues, the fiber tracts of the brain, convey the output of one capacity from its locale to the locales of other capacities. The mind has an organizational chart, indeed, and it is a chart of capacities and subcapacities, but it is at the same time a chart of mental organs that are specific physical tissues inside the skull.

Meynert and Wernicke were not just pluggers, too absorbed with biological and clinical detail to concern themselves with the overall structure of mind. Meynert published a textbook on psychiatry in 1884, in which the general idea of a neurophysiology of the mind was developed. Wernicke wrote a series of books and essays with the same aim, including in 1879 an essay on consciousness. In 1894 another of Brücke's students, Sigmund Exner, who was only slightly senior to Freud, wrote a speculative neuropsychology in much the same spirit. In several ways, Exner's book provided the framework for Freud's early thinking about the mind and the brain.

Brentano and the neurophysiologists agreed that psychology should have exact laws, and that the goal of psychology should be to find such laws. They disagreed about everything else, and for the most part Freud's views reflect those of Brücke and Meynert, not the view of Brentano. Brentano held that there are exact laws that refer only to the mental, and do not need to appeal to physical circumstances. Brücke and Meynert and Wernicke held that the exact laws concern physical properties or concern the relationship of physical features to mental capacities. The exact lesions that will incapacitate people to produce speech may not be known, just as the exact mass of hydrogen may not be known. But it is a perfectly general law that if all of Broca's area is destroyed, the capacity for speech will be lost. Brentano, unfortunately, had no laws of any interest to propose, and while his *Psychology from an Empirical Standpoint* contains lively criticism, when it turns to producing "results" from Brentano's method the product is deadly dull and nearly vacuous. Wernicke's accomplishment in producing a new psychophysical hypothesis correlative with a new analysis of the capacity for language stands in stark contrast to Brentano's rather lame effort. Any scientific reader of both Brentano's and Wernicke's work, and Freud was surely such a reader, could not have failed to notice the extraordinary difference in clarity, detail, and accomplishment in the positive parts of the two books, even if, as Freud came eventually to do, one disagreed with Wernicke's theory of language capacity.

Freud was reared to think that *psychology should be a neurophysiology of the mental* in which the explanation of capacities in terms of subcapacities proceeds in pace with the identification of parts of the brain essential for the component capacities, and the explanation of the component capacities eventually becomes a matter of physics and chemistry upon which all other properties supervene. That way of thinking about the project of psychology is one thread in contemporary cognitive science. Freud learned this way of thinking about psychology, but for two reasons it does not describe quite how he thought about the matter, even from early days in his professional career.

There is the problem of the contents of consciousness. Although it is true that the kidneys cleanse the blood, a materialist physiology need not give an account of the property of "cleansing" in general, because there is no such property. But one cannot say the same for the contents of consciousness, for the taste of pineapple, for the desire to have sexual relations with another, for the stomachache. The properties of each of us revealed immediately through consciousness seem real enough (indeed so real that we cannot bring ourselves *not* to believe in them), and the phrases that describe them cannot be dismissed as terms of convenience, useful but signifying nothing. A neurophysiology of the mental has a further obligation, and that is to explain what the contents of consciousness are and how they come about. Wernicke and others realized as much, even if they did not know how to provide such an explanation.

And, for Freud, like many other students of neurology of the time,[2] there is the further complexity that he did not quite believe Wernicke's localization schemes, nor was he sure that *any* localization scheme is possible for cognitive capacities. Nor was he quite sure of the contrary, which is why, over nearly fifty years, he often said one thing and then another about the place of thought.

Freud took his medical degree in 1881. For the next four years he worked in laboratories and hospitals in Vienna, until in 1885 he received a traveling scholarship that took him to Paris to study with Charcot, the great French neurologist. He won the scholarship in part through Brücke's lobbying, and it was in the way of compensation: Brücke had told Freud he had no prospects of an academic career. Returning from France, Freud again took up work in hospitals and clinics until, in 1887, he began private practice as a neurologist. Although he was no longer doing anatomical research, and after he began private practice had neither time nor morgues for research on the localization of cognitive functions, Freud remained fully informed of developments in mental physiology through the middle of the 1890s. In small ways he even contributed to those developments.

Freud's style of argument in the 1890s was framed by the empiricist scientific standards of John Stuart Mill (some of whose social essays he translated for Theodor Gomperz's German edition of Mill's works). In private, in his manuscripts and in his correspondence with his friend Wilhelm Fliess, Freud developed a broad, speculative conception of mind and of the enterprise of psychology. That conception can be found in his letters and manuscripts, especially around 1895. Its major statement is a document later entitled *Project for a Scientific Psychology;* it was evidently originally intended for publication, but Freud was uneasy with it, and seems to have submitted it to no one but Fliess. Late in his life Freud attempted unsuccessfully to have the manuscript destroyed. Commentators since have been struck by how much of the *Project* echoes through Freud's later work; we find pieces of its formulations in *The Interpretation of Dreams,* in "Instincts and Their Vicissitudes," in *The Ego and the Id,* in Freud's posthumous *Outline of Psycho-Analysis,* and we find its terminology throughout Freud's subsequent writings.

The *Project* really was Freud's project; it states the understanding both of mind and of the aims of psychology that governed his work in the 1890s, and that remained a part of his conception throughout his life. In major respects, Freud's conception was that of many cognitive psychologists of our own time. Once again, Freud was not singularly prescient; his perspective was shared by many of his teachers and colleagues, and his *Project* is largely an adaptation of their views. The similarity between Freud's enterprise and enterprises of our own day is less a cause for wonder than an aid in understanding both him and us.

I have argued that at least in the early part of his career, Freud conceived of himself as doing mental physiology, and that he shared the enterprise with many of the neuropsychologists of his day. The *Project for a Scientific Psychology* is his clearest and bravest attempt at a physiology of the mind. The most striking difference between that enterprise and contemporary cognitive science is that we possess the computer, and the computational pictures of how the mind works that the computer has provoked. To see the connections between what Freud was about and what contemporary cognitive psychologists are up to, we must consider the analogies between physiology, on the one hand, and computer science on the other. Freud aside, the analogies are essential to what cognitive science is supposed to be about. Once the analogies are briefly described, we will return to the details of Freud's mental physiology, and see how profoundly our novelties are echoes too.

II

Computing machines have an architecture or structure, just as the human body does. One can do a physiology of computers as well as (indeed more easily than) a physiology of the brain. Part of my digital computer is machinery for input and output; part of it is random access memory; a physically distinct part of it is memory storage; part of it is a central processing unit that performs operations in binary arithmetic; part of it is buses that connect the pieces. The different pieces of hardware have different functions, and can be functionally described, just as parts of my car can be, and parts of my body.

Computers have a physical structure, and the physical parts have functions. Without a program, those functions cannot be performed. In conventional computers a program is a set of instructions that is stored in the machine memory and then carried out, sequentially, when the computer is given an appropriate input. We usually specify the instructions in a "high-level language" such as PASCAL or LISP; in a proper machine, instructions written in such languages are translated into instructions that cause the physical parts of the machine to act appropriately. The program, the LISP code or the PASCAL code or the machine code into which it is translated, determines a sequence of computational stages for every possible input. The program determines a function from inputs to outputs, but because the sequence of computational stages may be infinite for some inputs, the function may not be defined on all possible inputs. The partial functions so determined are ipso facto computable functions. This way of looking at things enables us to ignore the physical details and consider simply the abstract structure of a method of specifying programs. Any such method, such as LISP or PASCAL, or a machine language code, is a *programming system*. Ideal programming systems permit the expression of programs for every computable function, and in fact an infinity of different programs in the same programming system will compute one and the same computable function.

There is an infinity of different programming systems that are

equivalent in defining programs that will compute exactly the same class of computable partial functions. Programming systems have a kind of formal or mathematical structure quite aside from any physical implementation. Each one of them represents a way of organizing computing, an "architecture," if you will. The study of the structure of programming systems is not computational physiology because the study of formal structure need not be concomitant with a study of physical structure. We can get a little closer to physiology if we consider the notion of a *machine model* which I, perhaps idiosyncratically, take to be the combination of a programming system and a story. The story says what *kinds* of physical pieces might realize the programming system. A universal Turing machine is a familiar machine model. There is a programming system, which could be given as a finite mathematical object, and there is a story about how the programming system might be realized. In the story, there is a tape with squares upon which elements of the input vocabulary may be written; there is a movable "head" that is always at one square or another and can read what is written on that square and can also write something else in its place; there is a machine table that contains "states" that tell the head what to write and how to move and determines the subsequent state. The Turing machine story does not describe any particular physical object, but it describes an imaginable *kind* of physical object with separate parts having specific computational functions and relevant capacities, and it connects that kind of physical object with a programming system. The result is that we can see how objects of that kind could carry out computations.[3]

A machine model is not a piece of computer physiology, but it is exactly the sort of *theory* we could use in doing computational physiology. If one wanted to understand how it is that a device one thinks might be a computer is indeed a computer, one would want to identify the physical parts of the object with the parts of a machine model and to show under that identification that the physical object goes through a sequence of states corresponding to the stages of the associated programming system. Identifying a physical object, or a class of physical objects, as instances of a machine model is clearly an inductive task; the identification represents an empirical claim, and evidence consists of observations of the internal and external behavior of objects in the class. Not only is it an empirical task to identify an actual physical object as a computer that realizes a particular computational model, in the worst case it is a daunting empirical task. The class of possible theories to be considered is enormous; there is an infinity of different programming systems, and the number of machine models is therefore bounded only by the possibility of telling physical stories to go with the programming systems. We can imagine Turing machines that have not just one but any number of tapes. We can imagine that there are addressable registers rather than tape squares. We can imagine physical processes, such as cellular automata, that are very remote from our usual notion of machinery, but that still represent machine models. Sometimes the story comes first, the programming system second; we may have a physical idea about how computation could be carried out without having a fully articulate formal understanding of the associated programming system. We may sometimes know what particular physical arrangements ought to compute without knowing quite how to classify things more generally. In science, intuition and theory play leapfrog.

Now the very idea of contemporary, computational cognitive psychology is that *we* realize some machine model or other; the goal of cognitive psychology is to do computational physiology *on us*. There may be no one thing that contemporary cognitive scientists believe, but there are characteristic theses. Cognitivists hold that the brain is a system that computes, and that its computations produce the phenomena of learning, perception, memory, language, imagination, and so forth. They begin to differ when one asks what sort of computer the brain is, and how and what exactly it computes. Some say that the brain is a *symbolic* computer, which sounds utterly redundant, since a computer that computed something other than symbols would be a factory. But they mean something more than that; they mean, at least, that the brain is a computer that encodes propositions and images in physical variables and states. The analogy is with machine states in a digital computer. Physical configurations in the machine encode propositions or imperatives that can be expressed in programming languages. Physical configurations in the brain encode propositions or imperatives or images that can be stated in English, or psychologese, or PASCAL, or can be depicted. The brain is a computer with a *language*, the language of thought.[4]

III

Many cognitive psychologists see the brain/computer as having a physical structure that is computationally relevant, and that realizes some programming system, in just the way that a real physical computer has a physical structure that is relevant to its computational functions. Of course they do not regard the brain as a computer organized in just the way IBM now designs them, but they do think of the brain as having specialized, physically distinct pieces that have particular causal and computational roles in producing various human capacities such as visual memory, or visual image formation, or speech recognition, and so on. They think of the brain as executing procedures, not necessarily serially. Sometimes a more or less explicit programming system is proposed by psychologists, but more often the suggestions are partial and fragmentary and focus on the functional roles of hypothetical pieces in some not yet fully explicit machine model. The theory of computation forms the theoretical backing for the enterprise of cognitive psychology, but the particulars of the formal theory are rarely used. Which is, in part, why contemporary work is so much like the enterprise of nineteenth-century neuropsychology. Freud and his contemporaries had no glimmer of the notion of a programming system, but they certainly thought of the brain as a biological machine that manipulates symbols, and they certainly thought that particular physical pieces or aspects of the brain have special roles in those manipulations. Although Freud could not have known it, his speculations about mental physiology are as much speculations about the machine model of mind as are the theories of our contemporaries. The differences between Freud's contemporaries and ours are largely in manner of speech, not in manner of thought. To see just how close the thoughts are, let us consider two contemporary approaches to the computational physiology of the mind.

There are two main contemporary views of the computational structure of the brain, although each view has many variants, and there are many attempts at compromise. Those who follow one main line in cognitive psychology regard the brain as executing instructions serially; the instructions, in turn, are stored somehow within it. There is another, apparently quite different, computational picture of the brain. The initial idea was to take more seriously the superficial anatomy of the brain, and to build machine models that have some faithfulness to it. The brain's structure is cellular, and the cells connect through the synaptic connections structure of the nerve cells. This suggests a network, or more precisely a graph, whose vertices are the cells and whose edges represent synaptic connections. Exactly this picture was suggested during the days of cybernetics by McCullough and Pitts. It has been revived in recent years under such titles as "Parallel Distributed Processing" or "Connectionist Machines." The network and the algorithms for modifying its characteristics can, if one insists, be viewed as a kind of fixed, hard-wired program, but the algorithms or instructions for such networks specify the behavior of individual network nodes and links more or less separately; each node or link executes the instructions pertinent to it alone.

A variety of connectionist devices have been proposed; one example will have to suffice. Consider a network in which each vertex can have only one of two states, *on* or *off*. Suppose, further, that every edge in the network has a numerical weight, either positive or negative, attached to it. Think of the state of each vertex as a random variable, and suppose that the probability at any moment that a particular vertex v is *on* depends *only* on the vertices adjacent to it that are *on* at the same moment, and the weights of the edges connecting those vertices with v. If we start such a network in some state, then the state will change over time, as vertices flash *on* and *off*. If we let the network run for a long while, there will be a long-run frequency with which any particular vertex is *on*, and there will therefore also be a long-run frequency with which each possible state of the system (that is, each possible assignment of values 0 or 1 to every vertex) occurs. So there will be a long-run or "equilibrium" probability distribution over the states of the system. Now it turns out that associated with any state of the system there is a function determined entirely by that state and the weights of the edges in the network, and that function looks formally very much like the energy function of statistical thermodynamics. The equilibrium probability distribution over the states of the network is in turn a function of the energies of the states. In fact, on simple assumptions, the equilibrium probability distribution looks like the Boltzmann distribution of statistical thermodynamics. Put simply, networks of this kind tend toward the lowest entropy states available to them.

Boltzmann machines can be made to learn. More accurately, procedures can be described that alter a Boltzmann machine until it computes some independently specified function. Boltzmann machines learn by a kind of analogue to facilitation in which future behavior is altered by the previous occasions in which the internal nodes of the system have been activated. In practice, Boltzmann machines learn very slowly. In addition to Boltzmann machines several other kinds of distributed processors, or connectionist machines, have been described, with a variety of different learning procedures.

Connectionists cite Karl Lashley and Donald Hebb as their sources. In the 1920s Lashley, an American-born-and-educated physiological psychologist, emphasized the holistic character of brain processing. Hebb, in 1939, suggested that learning takes place in the brain by facilitation, and in particular that the more frequently a neural pathway is activated the more probable it is that it will be activated on subsequent occasions. Lashley and Hebb no doubt deserve their credit, but contemporary connectionists would be more accurate if they traced their sources to Hermann Helmholtz, Sigmund Exner, and Sigmund Freud. While the algorithms will not be found in the writings of Freud and his contemporaries (nor in Lashley or Hebb, for that matter), all of the other elements of connectionism are there, including even the notion that analogues to thermodynamic principles govern the processes of the connection machine that is the brain, and the idea that learning takes places by neural facilitation. Freud himself anticipated both the views of Lashley and Hebb, and presented them in detail that is more congruent to current thinking. In 1891, in his book on aphasia, Freud embraced a holistic account of brain functioning that is essentially the same as Lashley's. By 1894 he had mixed that picture with the views, championed by Meynert, Wernicke, Lichtheim, and others, that the brain contains physically distinct processing modules. The result was theoretically of a piece with the kind of work we find published by many contemporary cognitive psychologists.

Freud and his contemporaries already knew enough of neural anatomy and physiology to make many of the same general guesses about how the brain computes that are made by our contemporaries.

In particular, exactly like the cognitivists of our day, Freud held the brain to be a machine, and although he did not use the word, a machine that computes, and whose computational processes explain our behavior and our experience. Further, like many of our contemporaries, Freud held there to be a private, innate language of thought in which propositions are expressed and which acts as the fundamental coding in the brain.

Freud's machine model was a collection of neurones joined together at synapses like the vertices of a graph. He held the computations of the system to be governed by quasi-thermodynamic principles, and in particular by the principle that the system seeks the lowest energy state. Again like many contemporary connectionists, Freud held that learning takes place by facilitation. And finally, we will not much misunderstand Freud's enterprise – not just in his secret *Project*, but also in *The Interpretation of Dreams, The Ego and the Id*, and elsewhere – if we take him to have been seeking a machine model of the mental functioning of the brain. In none of this, save in some of his hypotheses about the structure of that model, was Freud particularly original.

Freud's *Project* begins with these words:

> The intention is to furnish a psychology that shall be a natural science; that is, to represent psychical processes as quantitatively determinate states of specifiable material particles, thus making those processes perspicuous and free from contradiction. Two principal ideas are involved: [1] What distinguishes activity from rest is to be regarded as Q, subject to the general laws of motion. [2] The neurones are to be taken as the material particles. (1950a [1887–1902], I, 295)

The picture of the nervous system we obtain from Freud's *Project* goes roughly like this. The nerve cells are connected at synaptic junctions; they pass something among them that changes their physical energy state. Denote this something, whatever it may be, by "Q," for quantity. There are two ways in which Q might increase in the nervous system: through stimuli from the external world, and through "internal stimuli" from the cells of the body, which is to say through the internal chemical mechanisms of the instincts of hunger, thirst, sex, and so on. The amount of this quantity in the nervous system is not constant but can be increased or decreased by internal and external causes. The nervous system, as Freud conceives it, behaves like any other physical system; it tends to the lowest possible energy states, and the state transitions have a psychological correlate. Increase in energy, or Q, is painful, decrease is pleasurable. The organism is so structured that it reacts automatically to avoid the increase of Q from external stimuli by automatic motions, or reflexes. But Q from internal sources cannot be avoided by reflex motions. To shut off the internal sources of excitation requires rather definite physical situations and the motion of the organism must therefore be directed toward realizing them. The hungry baby, for example, must find the mother's breast. Freud supposed that such motions are carried out by a kind of computational process in which energy is stored up in the nerve cells temporarily. That store constitutes thought and desire and plan, and the nervous system tolerates it only because it leads, in the long run, to lower internal excitation than would otherwise occur. Freud calls the store of energy in a nerve cell "cathexis."[5] When a collection of nerve cells and their energy state represent the memory of a thought, Freud says the thought (or the "idea") is cathected.

Freud supposed that the cells of the nervous system are not all of one sort with regard to their changes of energy state. Some cells, he supposed, are unaltered by the passage of the unknown Q through them, while another class of cells is changed in a quasi-permanent way. The second class, the psi neurones, are responsible for memory, planning, goal-directed movement, and so on, but their processes are not *conscious*. They can have their energy states raised and kept raised; Freud says they are cathected. For Freud, learning is fundamentally adapting an energy distribution among the psi neurones, and it is accomplished by *facilitation* and cathexis. For example, if a is a nerve cell connected with cells b, c, and d, and a and b are cathected, then proportionately more Q passing though cell a will move to b than will move to c or to d. Moreover the passage of Q along any path is subject to a threshold; unless the difference in Q values is high enough, no Q will pass at all. So the cathexis of cells a and b inhibits passage of Q from cell a to cells c or d. If cell c is what Freud calls a "key" neurone, one that controls somatic cells generating Q, then because of the facilitation between a and b, the passage of Q through a is likely not to stimulate c; the facilitation between a and b prevents Q from increasing in the system.

This much of Freud's *Project* is in the same spirit as contemporary work on connectionist models of mind, and it is motivated by much the same picture of the mind and much the same level of anatomical and physiological detail. Connectionists propose that the brain is a computational network that functions to minimize entropy and that learns by facilitation. Freud has no algorithms, and his usage is not entirely consistent, but he says something analogous. *The economic viewpoint, the pleasure principle, really is Freud's computational model.*

Freud's general conception of connectionist learning is different from the framework of our contemporaries in one important respect. In that respect Freud's view is novel and deserves technical attention – attention that it will not be given here. Contemporary connectionist learning algorithms are essentially static; they modify a network to approximate a fixed probability measure. Freud's conception is more genuinely dynamic. The energy of the network is viewed as *potential energy* that the system tends to minimize; the network is not isolated but is instead subject to energy shocks. The energy shocks depend on the response the network gives to externally imposed inputs, and the effect of any shock is to add energy to the network. Freud thinks the system learns by adjusting weights (and more or less fixed *on* or *off* values for certain network nodes) that will tend to minimize the energy shocks in the long run. The network learns through psychological Darwinism; those network arrangements are fittest that minimize the long-run energy shocks, and the fittest survive. Essentially, the nervous system is represented as a subcomponent of a larger, constant energy system; energy transfers in and out of the subcomponent must occur through specific nodes. Energy inputs to the subcomponent are determined by some externally imposed schedule, and the problem is to find an algorithm for adjusting the subcomponent's weights on node links that will minimize the expected energy of the subcomponent for every externally imposed schedule. Just how the adjustment takes place Freud does not say. Freud's conception of how the nervous system learns is a kind of compromise between contemporary connectionist algorithms, of which the Boltzmann algorithm is one example, and contemporary "genetic" learning algorithms, that also use Darwinian ideas.[6]

Connectionist psychologists of our day sometimes want to superimpose upon their computational picture a notion of computation in which there is a language of thought; Freud did the same, although he did not write of languages but rather of "ideas." Freud supposed that a collection of cathected neurones constitutes a "memory image" of an object or circumstance. These memory images are the objects of propositional attitudes: They may be desired, or wished, or feared, or believed. Freud makes it clear that they have a *linguistic* structure. Thus when writing about "Cognition and Reproductive Thought" in his *Project* Freud says:

> Let us suppose that, quite generally, the wishful cathexis relates to neurone a + neurone b, and the perceptual cathexis to neurone a + c. Biological experience will teach here once again that it is unsafe to initiate discharge if the indications of reality do not confirm the whole complex but only a part of it. A way is now found, however, of completing the similarity into an identity. The perceptual complex, if it is compared with other perceptual complexes, can be dissected into a component portion, neurone a, which on the whole remains the same, and a second component portion, neurone b, which for the most part varies. Language will later apply the term *judgment* to this dissection and will discover the resemblance which in fact exists between the nucleus of the ego and the constant perceptual component and between the changing cathexes . . . [of desire]; it [language] will call neurone *a* the *thing* and neurone *b* its activity or attribute – in short its *predicate.* (327–8)

Freud had only subject and predicate, and none of our programming systems, but he most certainly had the notion of a language of thought. Moreover, it is perfectly clear that Freud regarded the language of thought as preceding all natural language and in a way independent of it. Thus babes have wishes, perceptions, and judgments whose content is represented in the language of thought even before they have the language of their mothers. So too, the representation of words and the representation of "ideas" are distinct, and one of the mechanisms for evading repression is, according to Freud, to bring an idea and a corresponding word or description in natural language into association.[7]

Freud's view is that we are biological machines; we compute and learn by means of the pleasure principle, and we change our state according to physical law. Our nervous states include energy distributions that are representational and have a linguistic structure that arises spontaneously, before any natural language is learned. Hear how Freud continues his theory of the mechanisms of wish and judgment, and how they produce motion:

> If neurone a coincides [in the two cathexes] but neurone c is perceived instead of neurone b, then the activity of the ego follows the connections of this neurone c and, by means of a current of Qn along these connections, causes new cathexes to emerge until access is found to the missing neurone b. As a rule, the image of a movement [a motor image] arises which is interpolated between neurone c and neurone b; and, when this image is freshly activated through a movement carried out really, the perception of neurone b, and at the same time, the identity that is being sought, are established. Let us suppose, for instance, that the mnemic image wished for is the image of the mother's breast and a front view of its nipple, and that the first perception is a side view of the same object, without the nipple. In the child's memory there is an experience, made by chance in the course of sucking, that with a particular head-movement the front image turns into the side image. The side image which is now seen leads to the head-movement; an experiment shows that its counterpart must be carried out, and the perception of the front view is achieved. (328)

To see how close Freud's conception is to contemporary views, or, if you prefer, to see how little we have progressed, it is useful to compare these passages with a contemporary discussion of distributed processing:

> The very simplest distributed scheme would represent the concept of onion and the concept of chimpanzee by alternative activity patterns over the very same set of units. It would then be hard to represent chimps and onions at the same time. This problem can be solved by using separate modules for each possible role of an item within a larger structure. Chimps, for example, are the "agent" of the liking and so a pattern representing chimps occupies the "agent" module and the pattern representing onions occupies the "patient" module.

The authors go on to give the following description:

> In this simplified scheme there are two different modules, one of which represents the agent and the other the patient. To incorporate the fact that chimpanzees like onions the pattern for chimpanzees in one module must be associated with the pattern for onions in the other module. Relationships other than "liking" can be implemented by having a third group of units whose pattern of activity represents the relationship.[8]

While Freud suggests that activation of individual neural states represents subjects and predicates, and a pattern of activation represents a judgment or wish, these contemporary connections instead suggest that patterns of activation among groups of neurones represent subjects and predicates. The differences are not large. In many other connectionist models, just as in Freud's model, individual nodes represent subject and predicate.

In Freud's *Project*, the infant is described more or less as an android run by a connectionist computer. If the details are a little hazy, and perhaps if we press even incoherent, still I think there is little doubt that Freud's conception of psychology and of the functioning of the mind is much the same as that of our contemporaries. I say again that there is not much new in it, and Freud is but a window to his time. Brücke and Wernicke had speculated, and so had Meynert, and in 1894, the year before the *Project* was written, Sigmund Exner, who had worked with Freud in Brücke's laboratory, published his *Entwurf zu einer physiologischen Erklärung der psychischen Erscheinungen*, which Freud's *Project* imitates in some detail. Of course Freud is original and peculiar in certain ways; between investigating belief and investigating desire, Freud always preferred desire, and his psychology is more a theory of wishing than of learning.

Freud's problems are our problems. Consider only the question of consciousness. The evident phenomenal fact is that consciousness is serial and in normal people unified. Freud's French contemporaries, and others taken by the phenomena of multiple personalities, were happy to hypothesize parallel consciousnesses in one and the same brain, but Freud did not. There is one unified consciousness, and in it one thing happens after another. We can recall not only what we have done, but in most circumstances the sequence of our actions. We view our own actions – at least our recent actions – as our own, not as the actions of a stranger. But Freud's machine model is not serial, it is a parallel distributed processing model in which there is no innate control unit, and nothing intrinsic to guarantee coordination. Each nerve cell does its thing, affected only by those cells that synapse with it. Thus for Freud the unconscious, or what he later called the id, is a collection of nerve cells with independent representations; as thoughts, the representations corresponding to the cells of the id may be inconsistent, they are not subject to logical processing, and they do not *occur* serially the way conscious thoughts do. Freud says the id is not subject to time, and he claims thereby to refute Kant. Freud's picture of the id is just the sort of thing we might naively expect from connectionist computation. It is just the sort of thing we do not find in consciousness. Somehow, if the connectionist picture is right, serial computation (or something that looks and feels like it) must emerge from the connections. Freud had no serious idea as to how, nor do we. His only suggestion is that consciousness is due to wave properties of the physical energy of the nerves, and that some nerves are specially equipped to detect the wave properties. The proposal is physically jejune, but even if we suppose it we obtain no explanation of the unity and serial character of consciousness.

Freud's conception of psychology in the middle of the 1890s is of a physiology of the mind in which the description of function, capacity, and physical structure and process are concomitant and inextricable. In the next decades Freud began to extricate them, and thus created a body of questions that apply as much to contemporary cognitive psychology as to psychoanalysis.

IV

Between 1885 and 1898, or thereabouts, Freud labored to stay abreast of developments in neuropsychology. Freud's book on aphasia, published in 1891, is evidence of that attempt. The private *Project* shows as much; its neurophysiology is up to date, and in many ways it simply copies the ideas of Sigmund Exner's *Entwurf*, which had appeared the year before. But in the long run Freud could not hope to continue making contributions to neuropsychology. He lacked both laboratory and morgue to do original work. Still, while he could leave neuropsychology, he could not leave the general conception of the mind and of psychological science upon which he had been reared. What he could do is separate and qualify its pieces, and he tried.

In physiology the analysis of function goes hand in hand with the identification of organic structures and the determination of their causes and effects on one another. In their different ways, Wernicke's work on aphasia and Freud's *Project for a Scientific Psychology* attempted to do the same thing for the mind. But when Freud turned to private practice he was confined to clinical evidence, to the evidence of his patient's behavior, their histories, their memories, their errors; he could not get at their brains. The result was that he began to attempt to characterize the functional structure of mind without a concomitant physical basis, without the organs of function (the ego, for example, or the dream censor) having any identification as specific tissues, without their causes and effects identified as specific kinds of physical changes.

So it happened that in the years after 1898, Freud often described mental processes and entities in terms of their *functional role:* in terms that is, of what they do to one another and to behavior, not in terms of physical characteristics. The mechanisms of defense, repression, the dream work, and later the id, the ego, and the superego are characterized by what they do to one another, and by how they together determine behavior.

Now in fact what I have just written is a half truth. It is half true that after 1898 Freud characterizes the mind functionally without concomitant physics. In fact, he is radically inconsistent, as though, depending upon your point of view, either he could not shake old bad habits, or he could not escape the fundamental soundness of his earlier physiological approach to the mind. Throughout the rest of his career, Freud explained behavior by appeal to the "libido," which in one reading is nothing other than his term for whatever part of the real physical psychic energy is due to sexual sources. In *The Interpretation of Dreams* there is a last chapter taken principally from the unpublished *Project.* Freud warns the reader that the elements of the theory are not to be assumed to have discrete and distinct physical locations, but he also makes it clear that the "systems" he describes and the processes among them are thought somehow to be realized in the brain by "neuronal excitations." In 1914, in his paper on the unconscious, Freud renounced a physiological significance for his theory "at least for the present." But he could not stay away from physiology and anatomy for long; much of his 1915 essay "Instincts and Their Vicissitudes" comes directly from the *Project*, and in the last decade and a half of his life he repeatedly gave his functional structures a physical locale. Thus in 1917, in the last chapters of his *Introductory Lectures*, Freud offered hypotheses about the physical location in the brain of various functions. *Beyond the Pleasure Principle*, published in 1920, was, like the *Three Essays on Sexuality* fifteen years earlier, a biological tract based on psychoanalytic evidence, and it made again many of the points made in the *Project,* and made them in the same language. Parts of this book, and passages in *The Ego and the Id* as well, are unintelligible unless we read Freud's theory as in part a theory of the physical partitioning of the brain's functions. In Freud's last works, *Moses and Monotheism* and *An Outline of Psycho-Analysis,* the anatomical localizations conjectured in the *Project* are again asserted.

So it seems fair to say that Freud thought he could characterize a *functional structure* for the mind without at the same time identifying the physical basis of that structure, that he thought the functional structure was somehow realized by the excitations of the brain cells, and that he could not keep himself from intermittent speculations about the physical locales of some of these functions. Cognitive psychologists nowadays attempt to describe the procedures by which cognitive capacities are exercised. Save for the cognitive neuropsychologists, they usually do so without much or any regard for the physical basis or locale of the procedures. Now and then an anatomical or physiological speculation will slip in. They have voluntarily embraced the separation of substance and function to which Freud was driven by necessity, and philosophers have made the separation into a metaphysic. Many psychologists, and philosophical commentators, avoid talking of machine models altogether, and prefer instead to claim their goal is the discovery of the "functional architecture" of mind. Of course, there is no harm in using different words, but the words are chosen to a point. The point is partly, I suspect, to avoid reference to the formal theory of computation, which many psychologists do not understand and do not much care about; but more important, the point is to emphasize the thought that the story that goes with a machine model is not, contrary to my usage, a story of *physical* kinds. In this view, the story given in a machine model does not describe a physical kind but instead describes something that is different in principle, a *functional* kind.

V

A homuncular explanation accounts for the actions of an agent by the actions of littler agents that compose it. Homuncular explanations have traditionally been despised on the grounds that they are

circular; they appeal not just to events that are as puzzling as the events to be explained, but worse, to events that are puzzling for the very same reasons as the events to be explained. If Judith's action in insulting Hermione is explained by postulating an entity within Judith that wished to insult Hermione and that makes Judith move, nothing is explained, at least not according to the philosophers.

Cognitive science has helped to make homuncular explanations seem more like genuine explanations. The very idea of functional analysis is to decompose capacities into relationships among subcapacities; if the means by which the subcapacities are effected remain for a while mysterious and the subcapacities can be described in terms of belief and desire, then for that while they can be thought of as homunculi. The decomposition is paralleled in the strategy of the computer programmer, who writes "big" functions initially in terms of names of slightly simpler functions, leaving for later a specification of those simpler components. Even with homuncular subcapacities, a functional analysis may enlighten us, contribute to our understanding, and do something explanatory.[9] Daniel Dennett says that homuncular explanations really explain provided the homunculi are stupider than is the agent whose actions they are to explain, stupider in that the homunculi have a more limited set of cognitive capacities than does the agent they compose.

Freud held a far more generous conception of the value of homuncular explanations, and I believe he was right to do so. In a sense, Freud's homunculi, at least some of them, can be smarter than the agent they compose, not stupider. Freud's conception of homuncular explanation derives from a more general strategy, namely to see the internal devices of the mind mirrored in the devices of social intercourse, in politics, in literature, in the theater. Freud grew to maturity in a time when Austria was in political and social turmoil; he had for a while liberal, even radical, political views, and took a keen interest in Viennese politics. His education was classical, and he maintained throughout his life a lively interest in the arts and their devices.[10] Those devices, made internal, became for Freud part of the strategems of mental representation.

Freud's views contain a kind of anticipation of the results of political and economic theories of our own time, and by transforming observations about collective decision making into a theory of mind, Freud created a homuncular theory that does genuinely – whether or not correctly – explain features of human action. More than that, Freud's theory provides the framework for *one* sort of explanation of a variety of phenomena that have concerned philosophy since Plato: actions that require an apparently paradoxical failure of will or reason, including self-deception, weakness of will, or acting against one's own better judgment, and weakness of reason or failing to consider in evidence or consequence what one knows to be relevant.

In the right contexts homuncular explanations genuinely explain. If we open Judith up and find within her a little person who through the magic of electronics causes Judith to move, and the little person tells us it wished to insult Hermione, we will conclude that the homuncular explanation was no pseudoexplanation at all, but a genuine and correct explanation. In this case, the right context is *physics;* Judith's interior is a piece of physics, and it is the physical and literal construal of the homuncular explanation of Judith's insult that makes the explanation explanatory. If the explanation were instead that there is no little man inside Judith, but rather Judith insulted Hermione because she was in a *functional* state *like* that of having a little man inside her who wished to insult Hermione, we might have a *real* pseudoexplanation. Construed literally and physically, the homuncular explanation is a real enough explanation, although not the sort we expect to be correct. Construed metaphorically, the homuncular explanation looks to be a pseudoexplanation for reasons like Molière's: it seems to say that Judith insulted Hermione because Judith was in an insulting-Hermione mental state. But are there cases besides little men in heads in which homuncular explanations genuinely explain and might even be reasonably regarded as correct?

Politics provides a context in which homuncular explanations are familiar, and their familiarity suggests that they provide some genuine satisfaction to the understanding. Some of the events in our world are events in which states do things, and governments take actions. How do we explain the actions of governments? Almost always, I think, in homuncular fashion. We explain the actions of governments through the beliefs and interests and desires and weaknesses of the people whom we say *compose* the government, and through the "functional" relations of those persons in their roles as parts of the government. We may even explain the actions of governments in terms of intermediate homunculi, such as coalitions or interest groups or corporations or the armed forces. We explain the actions of supernational bodies, such as the General Assembly of the United Nations, in terms of the beliefs and desires of homuncular agents that are governments. The popular press is full of such explanations, it invents them even when they are not appropriate: I am not arguing that a homuncular explanation is always the *best* explanation.

Homuncular explanations of the actions of a government or other social entity are especially useful when those actions taken together are irrational in the sense that an action taken to achieve one goal has that goal defeated by an action taken to achieve some other goal, and the incompatibility is part of the doctrine of the government, part of what it believes, or a trivial inference from its doctrine. That is commonly the case with governments, and explanations are therefore often sought. How do we explain the fact that the government of the United States, under the administration of Ronald Reagan, wished to reduce spending on social welfare including aid to dependent children, felt obligated to continue minimum support for indigent mothers and their children, yet reduced or eliminated abortion and birth control services for the poor, even while the government recognized that the absence of those services could only increase the numbers of children who required public support? The collection of beliefs and actions is puzzling because it is so palpably irrational, so straightforwardly *stupid.* No matter what consistent things you might desire, you would not do as Reagan's administration did. We give a homuncular explanation of the government's irrationality: The government acts in accord with the interests of different groups on different issues, even though the government knows that those interests and actions are logically and causally connected, and that the connections make for incompatibilities; one group dominates on one occasion and one issue, other groups on other occasions and issues. So we might say: Those who oppose birth control and abortion create sufficient political pressure[11] to undo government support for these activities; the middle class and the upper middle class, who for the most part favor or are indifferent to birth control and abortion, strongly favor a reduction in taxes and of the use of taxes to provide aid for the poor, and they create pressure upon the government to adopt such goals; everybody knows that sex causes pregnancy and pregnancy causes babies. Each of these groups *could be,* although I rather doubt they are, rational in the sense of having a consistent set of preferences. None need be diminished in its cognitive capacities in comparison with the government, although the government's *power* is greater.

Our time has made the irrationality of collective choice into mathematical theorems of various sorts. The original theorem was Arrow's.[12] The theorem says that under various technical assumptions, if there are at least two agents and three alternatives, then the only rule that will determine a consistent collective preference ordering of the three alternatives for every possible pair of preference orderings of the agents is a rule in which the collective preference ordering is, in every case, exactly the preference ordering of *one* of the agents. In understanding the theorem, the "rules" for determining collective choice need not be thought of as voting schemes; they can just as well be jousting tournaments or arm wrestling contests. Arrow's theorem is a result about political homunculi. If for the moment we think of rationality as requiring consistent preferences and nothing more, the theorem could be read this way: Unless one homunculus dominates in every possible case, an agent whose preferences are determined by the preferences of rational homunculi must, for some possible circumstances, be irrational.

Brentano taught Freud the doctrine of the unity of self. Freud did not believe it. According to Freud what produces action is not a unified self, but a collection of agents. The self is a collective fiction, like the government. The agents that compose a person have an identity through time and circumstance and they have a set of relations to one another; that identity and those relations, and nothing else, determines the identity of the person through time and circumstance. The homuncular agents differ in their desires and preferences. The actions of the person reveal a social choice, in something like Arrow's sense, determined from the preferences of the component agents by causes, by forces, rather than by voting procedures.

We know Freud's agents as the ego, the id, and the superego, but that classification appeared late in Freud's career, and is in any case too crude. Freud held the ego to be divided into a conscious and an unconscious part, which act in certain respects as agents with independent preferences. The conscious ego is rational and deliberate, something like the Mr. Spock of the society of the mind. It has detailed preferences about actions and thoughts. The unconscious ego has a funny set of preferences; it prefers to keep out of consciousness those thoughts that, were they to become conscious, would create enormous (conscious) pain. About everything else it is indifferent. The conscious ego, in a way, shares the preferences of the unconscious ego, but it cannot *think* them without agony, so (thanks to the unconscious ego) it does not think them. The id contains conflicting and inconsistent desires for the satisfaction of instincts, but it is indifferent to how those desires are fulfilled. The conscious ego cares a great deal about how, if at all, the id's desires are fulfilled, and so does the superego. The superego, the agent of conscience, has preferences over actions and thoughts, preferences more restrictive than those of the ego. Action results from the resolution of these conflicting preferences.

Freud's homunculi show many of the strategems of voters and voting blocks, and the life of the mind he assumes could, one thinks, be treated as a game of strategy played by several parties. Freud's agents try to conceal their preferences from one another; some agents censor the information that other agents attempt to send to one another. Freud's agents negotiate and make compromises and settle for their second and third choices when they cannot have their way. Of course, underneath all of this talk of agents and their wishes and compromises, Freud sees ultimately an entirely physical set of forces, compromising, if you will, by vector addition. Like a computer programmer, Freud starts with the big pieces, and tries to say what they do to one another, leaving as yet to be explained the mechanics by which they do it. The strategy is just the one Dennett describes, save that in an obvious sense Freud's homunculi need not be in the least stupider than the person they compose. If rationality is consistency of preference, then Freud's homunculi are more rational than persons. We may be equivocal, self-deceptive, suffer weakness of will, have inconsistent desires, but on Freud's account the homunculi within us need not.

I do not know whether Freud's homunculi are *necessary* to give a social explanation of individual irrationality, and the general question seems worthy of some attention. If an agent has an irrational (e.g., intransitive) set of preferences, what is the least number of rational homunculi into which he may be decomposed, such that the agent's preferences may be seen as collective preferences formed on the basis of the preferences of the homunculi? One would guess that in the absence of further constraints two homunculi suffice. If so, Pierre Janet's psychiatry, which explained neurosis by a "second consciousness," would seem more economical than Freud's. But of course the question may have more interesting answers if constraints are imposed on the preferences of the homunculi or on the rules by which the conflicting desires of homunculi may be accommodated.

Are Freud's homunculi physical or fictional or "functional"? The answer is a little equivocal. Most often, although certainly not always, Freud treats the ego, or at least the conscious ego, as a specific suborgan of the brain, usually the frontal cortex. The id is more vaguely characterized spatially, but Freud often writes as though it has some specific location. The unconscious ego lies between the two. The superego is characterized functionally rather than spatially. They are homunculi, but they are not *just* functional homunculi, they are (generally) also physical homunculi. Some of the homunculi, the ego for example, are *rational* agents, more rational than the person they compose. Even the id, if its conflicting preferences are regarded as the preferences of subhomunculi, could perhaps be thought of as a collection of rational agents. Or could it? What is required in order to gather together a group of desires and beliefs and call it an *agent?* What is going on when Freud separates our desires into the desires of distinct *agents* within us?

One story is that agency is what is required to explain and predict patterns of behavior, and there is nothing more to being an agent than exhibiting a pattern of behavior that can be explained by supposing there is a unified, more or less rational system of belief and desire.[13] On this view thermostats are agents quite as much as people, but it is not clear that Freud's homunculi will count. For the separate homunculi exhibit no "behavior" in the usual sense; all of their interactions are with one another, and the behavior of the individual they compose is not the behavior of any of the person's homunculi, but the effect of their negotiations and compromises. One might try somehow to extend the notion of behavior to include the goings-on internal to the mind, but within Freud's picture it would, I think, be a large undertaking to separate events that are explained as the actions of a single homunculus. More likely, we could extend the picture to something like this: To be an agent is to be a unified, more or less rational system of belief and desire that, *together with other agents,* explains a pattern of behavior. Some people would add then the system of beliefs and desires must be very large, and much like our own, but Freud would not.[14]

This does not explain what ties a collection of beliefs and desires together to make an agent. I cannot take one of your beliefs, one of mine, some of Saul Bellow's desires, and so on, and form a collection of beliefs and desires that is an agent. Why not? One insufficient reason is that the beliefs and desires are not localized in space, in the same head. Spatial distribution of beliefs and desires does not itself imply that the beliefs and desires are not those of one agent, as science fiction writers and philosophers both remind us.[15] In any case, the suggestion would only help Freud a little, since he is so equivocal about the existence of distinct spatial locations for his homunculi within the brain. A better explanation is that agency must bear a causal relation to action. A system of beliefs and desires taken from many people does not produce any actions; neither does it provide the reasons for any actions. The beliefs and desires of a normal, rational person both cause his action and provide reasons for it; not all beliefs and not all desires one has have a causal role in each action one undertakes, but virtually any belief and any desire are connected in forming possible reasons and possible causes for some potential action. In Freud's case none of the homuncular agents (save perhaps on some occasions the ego) are exclusively responsible for any action of the individual, and so this rather standard conception of agency does not straightforwardly apply. It does apply, more or less, if we socialize it. Roughly, what makes a system of beliefs and desires an agent is that they collaborate in almost every circumstance; they represent a vote in the society of mind, a society in which, to be sure, not all votes are equal. A collection of beliefs and desires forms a homuncular agent if the beliefs and desires are consistent and rationally combined to form preferences that are accommodated in the social determination of collective preferences and in the consequent determination of action by the whole individual.

Whether or not one believes in Freud's homunculi, Freud provides a *form* of explanation of action that is perfectly genuine, and might in appropriate applications even be correct. Freud's typical applications of his social theory of mind are to the explanation of irrational actions, especially the actions of neurotics, but the kind of explanation he provides also addresses ancient philosophical chestnuts.[16]

Reason and the will present puzzles that still feature large in the philosophy of mind. The puzzles concern familiar psychological phenomena whose reality we all recognize, but whose very description seems paradoxical.

We all recognize that people sometimes deceive themselves about their feelings, their desires, their reasons for action, even their beliefs. But self-deception seems to require that one and the same agent both know something and not know it at the same time, or both desire something and not desire something at the same time. And that seems not just unlikely, but *logically* impossible.

Ambivalence presents something of the same difficulty. Sometimes people seem to have analytically incompatible attitudes toward the same object. Their behavior rapidly alternates between animosity and affection toward the same person. We are inclined sometimes to say that a woman both loves and hates a man, or a man a woman. But to love is by its very meaning not to hate, and to hate is by its very meaning not to love, and so our common assessment of ambivalence seems inconsistent.

Weakness of will occurs when someone believes that, all things considered, a certain action is for the best, but succumbs to temptation and does not perform the action. With plausible assumptions the circumstance becomes paradoxical. Assume in addition only that agents want to do what they judge it best to do, and that if they do either of a pair of actions intentionally, they will do the action they want to do when they believe themselves free to do it, and we have a contradiction.[17]

There are weaknesses of reason that are at least as perplexing. Sometimes a person will sincerely want a certain outcome and sincerely believe that a certain action is necessary to obtain that outcome, and believe himself able to perform the action, and yet to all appearances deliberately fail to perform the action. Thus the infamous Professor Blondlot presumably knew what sort of experiments needed to be conducted in order to convince his contemporaries that his "N-Rays" were the real McCoy, but he did not conduct them, even though, historians seem to say, Blondlot was no mountebank. Sometimes a person will have evidence relevant to a conclusion, know it is relevant, and yet fail to use it, and draw an erroneous conclusion. Sometimes a person will know that a proposition is a consequence of what is believed, and yet fail to believe the consequence or to revise the beliefs of which it is a consequence.

It may be that not all of these difficulties are distinct, and that there is a reduction or commonality of pattern or explanation. Whatever the case, moral philosophy, and more lately philosophical psychology, have been concerned to explain these perplexities, or to explain them away, to show how they are possible, and why they are sometimes actual. It is straightforward to remove the apparent paradox in one or another of these cases by supposing the situation has in some way been misrepresented. For example, when someone has evidence that P is not the case, and knows it is evidence, and then ignores the evidence and asserts that P is the case, one need not be believing that what one believes to be disconfirmed is confirmed. We might instead explain the action by a kind of inward decision theory: The agent will choose to believe P or not according to which action has the greatest expected utility; believing P brings satisfactions if P is true, less satisfaction if P is false, but even though P is less probable than not, the expected utility of believing P is greater than the expected utility of not believing P. Pascal understood this sort of thing.

For Freud failures of rationality, or apparent failures, were the keys to the structure of mind, just as failures of speech were to Wernicke the keys to the functional structure of the brain. The interesting thing about Freud's social theory of mind is that it provides a mechanism for explaining not just one, but all of these paradoxes of will and reason. Moreover, the explanation is so obvious as to be almost irresistible, although not, I think, logically inevitable and certainly not necessarily complete. Freud did not seriously claim that his mode of explanation is exhaustive, and that such phenomena cannot arise in other ways.

A Freudian explanation of self-deception turns on the fact that the self is a collection P of agents, that what is known to one of these agents may not be known to another of them, and what is desired by one may not be desired by others, or be any of the desires attributable to the individual as a whole. What the id knows the conscious ego does not; what the id wants, the ego may not; what you want may not be what your id wants or what your ego wants. Any explanation of self-deception that supposes that we are composed of separate memory stores and that thought can occur while drawing from some of these stores but not from others, will be a Freudian explanation in spirit, whether or not the separate stores have the particular features Freud postulated. Sometimes accounts of this sort seem entirely plausible as an account of the phenomena of self-deception. A Freudian explanation of certain weaknesses of reason is of the same form. How is it that someone can neglect to consider evidence that is relevant to a conclusion, evidence that the agent knows about and whose relevance is also known, and evidence of a kind the agent is competent to evaluate? Easily enough if the agent has separate memory stores, and some of those stores are or can be made to be inaccessible to ratiocination. Freud's original examples are unconscious memories, but he expanded the framework, and the applicability of the explanatory strategy, to include the "preconscious."

Ambivalence is explained by supposing multiple agents with reasonably fixed but contrary preferences, and by supposing that no one of the agents always dominates. Freud's explanation of ambivalence in the Rat Man case goes like this: Conscious love and conscious hatred of one and the same object are possible provided neither is intense. When both become sufficiently intense they are incompatible and one emotion must become unconscious, generally the more painful emotion. Perhaps Freud can be understood as follows. One and the same agent cannot both love an object and hate that same object at the same time. But one agent can love *aspects* of an object and hate other *aspects* of an object. When attitudes toward aspects of objects become sufficiently intense, they become detached. They become attitudes toward the objects, not just toward aspects of the objects, and they therefore become incompatible. The rejected attitude becomes the attitude of some other agent within the self and helps determine the preferences of that agent. When the ego loves what the id hates there will be inconsistent preferences each of which will be revealed in varying circumstances, and there will also be sometimes a kind of indecisiveness. The phenomena of ambivalence are accounted for.

Weakness of the will is no more than ambivalence in action. One agent's reasons may be causes, but not reasons, for another agent.[18] One agent may decide that, all things considered, it is best not to have a further drink; the preference of another agent may intervene, and the drink taken. If one of the agents gives reasons and expresses

regrets, while the other is silent, we say the person was impulsive, that he gave in to temptation, that he had a weak will. Acts of incontinence betray an irrational whole that emerges from parts, homunculi, that may be more rational.

These are the ways Freud goes about explaining irrationality. His explanations may or may not be correct, but they are surely *explanations*. If that is doubted, consider that in each of the kinds of cases considered, whether ambivalence, weakness of will, self-deception, or weaknesses or reason, there are analogous phenomena in public life, and we routinely and sometimes correctly give Freudian explanations of these phenomena when they appear in the actions of governments, corporations, and other social entities. In the case of governments we know the homunculi exist, and who they are, and we can more directly verify the explanation offered. Freud's explanations of the self are less secure; they are not less genuine.

VI

Showing and saying have always been deeply entangled enterprises that somehow reach similar ends by disparate means. Saying has linguistic structure, logical structure, grammar; showing, to all appearances, has not. Showing is saying without chains. Every now and then there is an attempt to reduce one of the pair, saying and showing, to the other, or to establish the primacy of one to the exclusion of the other. In the early part of this century Wittgenstein, and the logical atomist movement generally, sought to reduce saying to a kind of showing. Later an heir of the movement, Nelson Goodman, sought to explain showing as a kind of saying. Several recent essays attempt to show the primacy of saying in the life of the mind, and psychologists continue to debate the autonomy of showing in mental life. Showing is certainly a way of saying, but since it lacks grammar and its objects lack grammatical categories, showing does not permit us our usual analyses of what is said. For most pieces of language we can give accounts of how they contribute to the truth value of sentences in which they occur; we do so by giving truth definitions that make the truth or falsity of sentences functions of the semantic properties of their component pieces. With pictures, with illustrations, with bits of theater, we can do no such thing. There are parts and uses of language that behave more like pictures than like sentences, and exactly this feature makes them puzzling and challenging for philosophical analysis. Demonstratives, thises and thats, can be used to show by saying, and for that reason they resist analysis by truth definitions. Metaphors and similes are refractory in the same way, and for the same cause; they are ways of asserting a showing.[19]

For Freud, who took his hypothetical forms of mental representation as much from the arts as from logic, the homunculi communicate both by image and by language, both by saying and by showing. Freud's accounts of the battles of the ego and the id and the superego read like little internal melodramas, and they are. The theater, above all art forms, is the place in which a complex thought can be both illustrated and said. Yet for Freud the theater of the mind is a kind of puppet show, controlled by purely physical forces that carry out computations; the show is the manifestation of the computations. Which brings us, implausibly, to Freud's views of the relations between computation and mental representation, and how the mind can work both by showing and by saying.

Connecting the *Project* with Freud's *Interpretation of Dreams,* published only four years later, we can extract a view about analog computation that bears on contemporary debates. The exercise has a certain ahistorical character, but historians of philosophy do not hesitate to offer Aristotelian, or Humian, or Leibnizian treatments of contemporary philosophical issues; I see no reason not to do the same for Freud.

Early in his career Freud, along with Breuer, thought of the symptoms of neurotics as a kind of aberrant reflex. Freud taught that behavior that seems aberrant and without rational structure may often have such a structure nonetheless, even if it is not evident. Freud's examples often concern the behavior of psychoneurotics. Thus his patient Dora, for example, will not give voice to the thought that she wants a family friend, Herr K., to make love to her, but Freud thinks she says it by playing with her reticule, and by her loss of speech when Herr K. is away. The actions are not speech, but Freud takes them to express a thought, usually by constituting an instance of the thought, or by being a little allegory. It is the same with Freud for internal actions as for external actions, for thoughts as for behavior. Dreams often seem to have no rational structure, but Freud insists that underneath, they do. The dream is usually an image or a sequence of images, proceeding as an inner theater of the absurd. But each play has, according to Freud, a message that it does not say explicitly but shows instead. The showing may be by pun, or by showing the opposite, or by excessive literalism, or by any of the other tricks of the theater. A woman in love with a conductor whom she regards as a towering figure dreams of a conductor in a tower above her.

The deepest novelty of *The Interpretation of Dreams* is the thought that literary and theatrical devices for representing meaning – the devices of parody, allegory, irony, exhibition, and depiction – may also be internal devices used in mental representation. The fundamental semantic insight is that the categories of proof and model theory are not mutually exclusive. One can imagine systems of expression in which some things are *said* by being *modeled,* and even systems in which things are said partly syntactically and partly by being modeled. In a way, the idea is easy and familiar. Almost everyone has seen children's books written partly in words and partly in pictures, with the pictures inserted in a line in place of a word or phrase, or sometimes in place of a syllable. Freud's thought is that mental representation works in a roughly similar way, in combination, of course, with irony and other devices.

If the difference between analog and digital computers is roughly the difference between proof relations and model relations, as I suggest, then one observation follows, an observation that might in any case be given other grounds: The class of computers cannot be partitioned into analog and digital. A computer can be both, or have features of both. A digital computer can be used to produce images, and the images can be used in analog computation. In principle, the analog output could be used to cause the input to another digital process, and so on.

Our usual formal systems, logics, make us think of accounts of inference as specifications of rules. Reasoning, ideally, is producing a sequence of sentences in accord with the rules. Syntactic rules permit the derivation of assertions based on the combinatorial properties of their syntactic components. There are notions of "semantic rule" in the philosophical literature, but they do nothing quite like what syntactic rules do. "Semantic rules" are usually, depending on the philosopher, either very general axioms (e.g., 'Everything colored is extended') or metalanguage statements about the *interpretation* of syntactic components. They are not analog inference rules. But I think we can imagine a *system of inference* that mixes proof theory and model theory, and contains analog rules of inference. Tracing out the derivation of a conclusion in such a system would amount to giving reasons for the conclusion, and some of the reasons would correspond to analog computations.

Our usual rules of inference for formal systems are combinatorial. Analog rules of inference cannot be. They must instead state general features of models that can be inferred to be features of the things modeled. We can imagine a language for talking of observable objects in the night sky. Let the language have the usual form of the predicate calculus, but let pictures of the sun, moon, shooting stars, comets, planets, and fixed stars also serve as individual names. Let the language be sufficiently interpreted that certain monadic predicates signify color terms: red, yellow, blue, and so on. Let the pictures come in various colors and suppose we add to the language the rule:

> *From any well-formed formula S, if p is a picture symbol occurring in S, and p has color r and R is a color predicate interpreted as r, infer S & R(p).*

In a system of inference that mixes proof and model theory, one can infer that the moon is yellow from premises that contain no color predicates but instead contain a depiction of the moon. (That color is modeled by color is of course irrelevant to the philosophical point.) An automaton that used such a system of inference would do some analog processing, and yet its conclusions about the colors of objects in the night sky would be "cognitively penetrable" in the sense that the processing would provide reasons for the conclusions. Perish the thought that there could be no such automaton, since something noncombinatorial must be done to apply the rule, namely it must be determined that p has color r. The detection of color can be done mechanically, as with spectroscopes, and our automaton can carry out derivations that accord with the rules of the system provided the automaton has some device for determining such physical properties of its representations. No homunculus is necessary for analog computation, any more than for digital computation.

One might object that in such an automaton the workings of the spectroscope would not be reasons, and that is so. The workings of

the spectroscope would *cause* certain representations and certain inferences to occur, but they would not themselves be reasons. And yet the workings could be woven into a process of inference so centrally that physical features of the spectroscopic process – such as the time it takes – become physical features of the reasoning process. More important, the physical output of the spectroscope could affect inference in a way that is cognitively penetrable. If, for example, what is inferred is a probability (e.g., of yellow) function of features of the measured spectrum, then that probability could be combined with prior probabilities in standard ways; the resulting inference to the conclusion that something is yellow will be determined both by the physical measurements and prior beliefs.

There is no difference in the philosophical point if the spectroscope is inside an automaton's head or in a physical laboratory. When a physicist looks at a spectrum, physical features of the spectrum combine with the physicist's prior beliefs to lead to a conclusion about the color of some object. Ordinary perception is a process in which "analog" features interact with digital features to produce reasoning; we have done no more than imagine that some of the analog features are themselves in the head.

The moral of the argument is that we can conceive of analog computation that, given an appropriate interpretation, forms part of a system of reasons for conclusions. A corollary, obvious in its own right, is that pieces of analog computation within a system that simulates rational behavior do not require special homunculi, and need not introduce special mysteries. I suppose the corollary has some practical bearing on disputes over mental imagery, but I do not mean to propose that our brains do actually implement analog inference rules of the sort I have considered. It would be charming if Freud were right after all, and if we worked by a mixture of syntactic representations and models, mixing digital and analog computation in our reasoning, but for all I know that may be altogether the wrong way to look at ourselves.

NOTES

[1] J. Kihlstrom, "The Cognitive Unconscious," *Science* 257 (1987): 1445–52.

[2] For example, Paul Mobius and Hughlings Jackson.

3 Compare J. Hopcroft and J. Ullman, *Introduction to Automata Theory, Languages and Computation* (Reading, Mass.: Addison-Wesley, 1979).

4 Compare J. Fodor's *The Language of Thought* (New York: Crowell, 1979). Fodor maintains that the brain has an innate, unconscious, utterly private *language*, a machine code if you will, in which thought finds expression.

5 The common translation from the German *besetzen.* Freud took the term and the idea from T. Meynert's *Psychiatry* (1884).

6 See J. Holland, *Adaptation in Natural and Artificial Systems* (Ann Arbor: University of Michigan Press, 1975).

7 Colin McGinn in *The Character of Mind* (Oxford: Oxford University Press, 1982) objects to the very idea of a language of thought that what is expressed in language may be expressed *insincerely,* and whatever the sort of "language" for thought supposed by cognitivists, it does not include insincere expression. While cognitivists in general can safely ignore this rebuff, it does not apply to Freud at all.

8 G. Hinton, J. McClelland, and D. Rumelhart, "Distributed Representations," in D. Rumelhart, J. McClelland, et al., *Parallel Distributed Processing,* vol. 1 (Cambridge, Mass.: MIT Press, 1986), pp. 82–3. Anyone who doubts the claim that much of contemporary connectionist cognitive psychology is reasonably viewed as nineteenth-century neuropsychological explanation plus the computer would do well to compare this volume with Exner's book and Freud's *Project.*

9 The best description of functional analysis is in R. Cummins, *The Nature of Psychological Explanation* (Cambridge, Mass.: MIT Press, 1983), but an earlier, vivid statement of the idea and the connection with homuncular explanation is to be found in D. Dennett's *Brainstorms* (Cambridge, Mass.: Bradford Books, 1978).

10 It is probably no accident that in the late 1890s plays about the unconscious meanings of dreams appeared in Vienna. For a discussion of the political background of Freud's youth, see W. McGrath, *Freud's Discovery of Psychoanalysis* (Ithaca, N.Y.: Cornell University Press, 1986).

11 Note how much the idiom is like Freud's, who speaks similarly of the "pressure" of instincts, or the "pressure" of repression.

12 K. Arrow, *Social Choice and Individual Values,* 2d ed. (New York: Wiley, 1963).

13 This view of agency is, I think, central to D. Dennett's *The Intentional Stance* (Cambridge, Mass.: MIT Press, 1987).

14 Compare Richard Rorty's "Freud and Moral Reflection," in J. Smith and W. Kerrigan, eds., *Pragmatism's Freud: The Moral Disposition* (Baltimore: Johns Hopkins Press, 1986).

15 See Dennett, "Where Am I?" in *Brainstorms.*

16 For an entirely contrary assessment whose arguments I find unpersuasive, see Irving Thalberg's "Freud's Anatomies of the Self," in J. Hopkins and R. Wollheim, eds., *Philosophical Essays on Freud* (Cambridge: Cambridge University Press, 1982).

17 These conditions are a paraphrase from Donald Davidson's "How Is Weakness of the Will Possible?," in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980). For the second conjunct to be plausible, "believe themselves free to" must be read as "believe themselves able to."

18 See Donald Davidson's insightful "Paradoxes of Irrationality," in Hopkins and Wollheim, *Philosophical Essays on Freud.* Save for the phrasing in terms of homunculi, my account of Freud's treatment of irrational action means to be in accord with Davidson's. Compare also D. Pears, "Motivated Irrationality" in the same place, and his book *Motivated Irrationality* (Oxford: Oxford University Press, 1984).

19 Compare N. Goodman, *Languages of Art,* 2d ed., (Indianapolis: Hackett Publishing, 1976); D. Kaplan, "D-That" in P. Cole, ed., *Syntax and Semantics,* vol. 9: *Pragmatics* (New York: Academic Press, 1978), pp. 221–43; and P. Machamer, "Problems of Knowledge Representation: Propositions, Procedures and Images," preprint, University of Pittsburgh.