For John Heil & Al Mele, eds.
Mental Causation
Oxford University Press
Ruth Garrett Millikan
University of Connecticut

Explanation in Biopsychology

I would like to explore implications for the science of psychology of the thesis that the categories of intentional psychology are function categories in the biologist's sense of "function", taking this to be a sense in which function is determined by evolutionary history rather than by current dispositions. I would like to explore, first, the general shape of the discipline that is psychology under this interpretation. What is its subject matter? What kinds of explanations does it seek for what kinds of phenomena? Second, I would like to bring these reflections to bear on the classic question concerning in what way, if any, does giving an individual's reasons for action tell of the causes of that action? I will not attempt to defend the thesis that the categories of intentional psychology are biological function categories, nor the thesis that biological function categories are carved out by reference to evolutionary history, but some clarifications of these theses, will of course, be in order.[1]

I.Teleo-functions

To describe the biological function of an item is not to describe its dispositional capacities. It is to describe the role that its ancestors played in a particular historical process, a concrete cyclical process of birth, development and reproduction extended over a large number of previous generations. It is to tell how earlier items involved in this historic process that are homologous to this functional item characteristically contributed to continuation of the cycle (thus helping, of course, to account for this item's existence). To say this is not to define the phenomenon of biological function. That can only be done, probably, by reference also to natural selection (see Millikan 1984, 1989b, 1989c). But I wish to call attention to the fact that the focus of the biological notion of function is on only very restricted aspects of the functional item's capacities, namely, those that have contributed over and over, in the same sort of way, to the historic cycle or chain of life.

Not every biological function of every biological item (type) is realized in every historic instance, say, in every generation. Some biological functions are very seldom performed. Still, they must occasionally have been performed, and performed in such circumstances as to weld an essential link in the historic chain of life, or they are not true functions. Consider, for example, the ability human babies are rumored to have of instant hibernation when submerged in very cold water. Surely it is a rare baby whose life has been saved by this capacity. But there have been enough, apparently, to fix the relevant genes in the gene-pool, hence to confer a biological function on this disposition. Alternatively, if the disposition should, as a matter of fact, have no such felicitous history, but arose only as a concommitant of other functions, as a "spandrel" (see Gould and Lewontin 1979), then it has no biological function. We should also note that not every functional item actually has the dispositional capacities to perform its biological functions. Homology is not identity; members of biological function categories can be malformed, diseased, or injured (see Millikan 1984, chapter 1). We can sum these points up by saying that biological functions are "teleo-functions" rather that "mechano-functions." They are biological purposes rather than

[1] These theses are defended in (Millikan 1984, 1986a, 1989a, 1989b, 1989c, 1990a, 1991a, 1991b).

activities or dispositions.[2]

II.Psychological Classification as Functional Classification; Categorial vs. Relational Functions

The claim that psychological classification is biological classification hence proceeds by reference to teleo-function includes that categories such as belief, desire, memory, percept, and purposive behavior are biological function categories--very broad and general ones, of course. Compare the categories limb, hormone, circulatory system, eye, visual system, etc.. More contentious, the claim includes that such categories or types as belief-that-it-is-raining, desire-to-visit-Paris, percept-of-a-cat, and purposeful-shooting-of-a-rabbit are carved out with reference to biological functions (though in the case of beliefs, not directly according to function; see section 9 below.) This more contentious claim presupposes two points that are not wholly familiar from a layman's understanding of biology.

The first point is that heredity does not directly dictate traits but rather patterns of interaction with the environment, thus controlling development. These interaction patterns control development not only before birth but also throughout life, so that how and what one learns is as much (and as little) dictated by heredity as is one's height and hair color. The second point is that the homologies among items that have historically played the same biological role in a species, and the homologies among the biological roles or functions that these homologous items have performed, are often highly relational. Homologous items may differ greatly from one another, as non-relationally or categorially described, both in structure and in function, their biologically significant similarities being captured only by multiply relational characterizations. To have biological functions an item need neither have the same categorial properties, e.g., the same absolute structure, as items that participated in the life cycles of ancestors, nor need its functions, when categorially described, be functions performed by any of its ancestors. Let me try to explain this clearly, for it is crucial.

Consider the neurological mechanism responsible for imprinting in ducklings. It has the relational function of imprinting on the duckling the visual character of something related to the duckling in a certain way, namely, as that which bears the relation mother of to the duckling. This relational function translates into a different categorial function [3] for the individual imprinting system of each individual duckling, since the visual character of each duckling's mother is different. Taking a still simpler example, this time from the domain of artefacts,[4] consider the function of a copying machine. It's relational function is to produce something that matches whatever pattern is put into its feeder. But given something particular put into its feeder, it then has the categorial function of producing a particular pattern. Notice that it is possible that this precise categorial function is one that neither it nor any other copying machine in history has ever happened to have before.

Similarly, humans are born with the capacity to develop concepts in accordance with certain general principles that operate upon the matter of the individual's particular experiences, and we are born with the capacity, in accordance with further general principles operating upon experience, to proceed to form desires and beliefs employing these concepts. These capacities are, in the first instance, relationally described capacities, but given the particular experiences of a particular individual, the biological functions of that individual's concept-forming

---

[2] What I am here calling "teleofunctions" I called "proper functions" in (Millikan 1984) and in most earlier papers.

[3] In (Millikan 1984) I called such categorical functions "adapted" proper functions.

[4] In (Millikan 1984) I argued that all teleology can be analyzed as belonging to the causal order in accordance with patterns analogous to those that establish biological teleology in the causal order. Therefore I use examples freely from the domain of artifacts as well as biology.

systems and belief-and-desire-forming systems translate into categorials. Likewise, the modifications of the nervous system that result, the instantiations of particular concepts, beliefs, and desires, have functional descriptions that are categorial. Further, these may, in many cases, be unique in history. Recall the individual mother-memory of the individual duckling, the function of which is to enable the duckling to recognize its individual mother, say, Sabatha. If the duckling has no elder siblings, then perhaps no biological device has ever had just that biological function before. This is the manner in which we speculate that belief-that-it-is-raining and desire-to-visit-Paris are distinct types carved out, in the end, by teleo-functional analysis.[5]

III. Biopsychology as a Study of Norms; The Ubiquity of Cognitive Failure

If the central categories of intentional psychology are indeed teleo-functional categories, this suggests that the core of the science of psychology should be a study of teleo-function. This core of psychology concerns the functions of the mechanisms that regulate those life-processes, those links in the life-chain from generation to generation, that are completed through the mediation of behaviors. We can call this discipline "biopsychology". The central concern of biopsychology is not to discover laws, neither universal laws nor statistical laws. Indeed, with a few rather special exceptions, the biological sciences do not typically traffic in laws. They seek to understand mechanisms that contribute to the cyclical processes that constitute development, maintenance and reproduction for the various species. But the rate of failure for many of these mechanisms is exceedingly high, especially when heavy interaction with the environment is involved.

These mechanisms unfold in anything but a lawlike manner. Biological functions are not, in general, reliable functions. They quite standardly go awry. Were this not the case, the world would be a marvelously populous place. The central job of the biological sciences is to describe biological norms, normative norms, not necessities or statistical averages. Indeed, these norms might better be called "ideals".[6] Let me detail this point as it applies, in particular, to the study of cognitive functions.

Some biological devices are such that a failure to perform their functions is immediately disastrous for the organism itself, or disastrous to its reproductive prospects. That is how it is, for example, if the heart or the kidneys fail. Equally often, however, functional failure is neither fatal nor the least bit dangerous. This may be for any of a variety of reasons. Some devices routinely get second chances, even multitudinous further chances, to perform their functions. Consider devices that regulate the performance of mating displays. Their teleo-function is to produce a display that will attract a mate, but if they fail on one occasion, they often have a chance to succeed on the next. Similarly, for predatory animals, the mechanisms that issue in food-procuring behaviors characteristically fail numerous times for every success. Many devices have functions that are redundant. They coexist with other devices that serve exactly the same functions in other ways. Thus the human system has several mechanisms redundantly devoted to cooling. These produce sweating, dilation of the capillaries, lethargy, motivate cooling-off behaviors, motivate the seeking of cooler spots, etc.. Similarly, many animals have alternative means of procuring food, so that if one fails, another may succeed. Alternative ways of doing the same often take the form of mechanisms that back one another up. For example, most animals possess reflex mechanisms designed to lessen the likelihood of physical harm (ducking reflexes, fall-checking reflexes), as well as exhibiting more sophisticated behaviors with the same purpose, but they also have mechanisms for repairing

---

[5] For fuller discussion of these issues, see (Millikan 1984, 1990a).

[6] In (Millikan 1984) I capitalized the "N" in "Normal" to remind that these norms are ideals rather than averages.

physical harm should it occur. They have mechanisms for preventing the entry of noxious bacteria, viruses, etc. into the tissues, but they also have mechanisms for destroying those noxious elements that do enter--mechanisms that themselves operate on a number of levels and with considerable redundancy. The multiplicity of devices aimed at the same end attests, in these cases, to the likelihood for each that it may fail.

We should be especially ready to expect failures in the case of mechanisms, such as the cognitive systems, that help to produce behaviors. This is because in order for behaviors to serve their biological functions, hence to complete the functions of the mechanisms that regulate them, mediation by the environment is required. Biological processes, portions of the biological cycle, that behaviors initiate are processes that loop through the world outside the organism. And it is obviously a great deal more difficult for the organism to stabilize its outer environment so as to provide the necessary conditions for completion of such loops than to stabilize its inner environment so as to complete, say, its physiological functions. So there is good reason to speculate that the cognitive systems might be abundantly unreliable in the performance, at least, of one portion of their functions--that portion which, unlike, say, inference and memory retrieval (though these are surely fallible enough), is accomplished through the mediation of structures and conditions in the world outside. Common examples of such failures result, we may suppose, in the acquisition of empty or confused concepts, in acts of misidentification of objects, kinds, stuffs or properties, in the fixation of false beliefs, in the acquisition of harmful desires, and in the failure of healthy desires to become realized. For each of these mishaps may be occasioned by failure of the environment to provide the ideal conditions which, unfortunately, are also necessary, for proper functioning of the cognitive systems. (Frequently it is, as it were, the world that fails us, rather than our inner systems.)

It follows that a description of the biological functions of the cognitive systems will in no way resemble a catalogue of psychological laws. It is certainly no psychological law, for example, that our beliefs are true, though it is a (teleo-)function of our belief-fixing systems to fix true beliefs.

Of course there are many biological functions that do get performed with pretty lawlike regularity, such as blood circulation and eyeblink reflexes, but it is not because of their lawlike properties that these functions are of interest to the biopsychologist. Turning the coin over, the frequency and, for the most part, the harmlessness of the occurrence of false beliefs, mistaken identifications, and so forth, should not cause us to suppose that these occurrences are biologically normal. Such failures may be frequent, conceivably they are even average, but they are not biologically normal. They do not exemplify patterns that have helped to forge links in the historic life-chain. Compare: being eaten by a bigger fish is the average thing that happens to little fish, but it is not on that account a biologically normal happening, relative to the little fish, nor is how little fish get eaten, as opposed to how they avoid getting eaten, a part of the ethology of little fish.

IV. The Subject Matter of Biopsychology is a Process

The biopsychologist is not like a physicist or, say, a mineralogist. The object of biopsychological study is not a chunk of matter, warm or cold, lying on the lab table waiting for its structure to be examined, for its input-output dispositions to be tabulated, or waiting to see what causes applied will produce what effects, what "special science" laws may hold for it. Nor is the point of biopsychology to examine or speculate on details of the complex structures inside the black box, to check on the dispositions of the components, nor to examine how the little dispositions inside add up to the complex dispositions of the whole. Biopsychology is not, then, all of what has traditionally been labeled "psychology". There are many industrial psychologists, for example, and many psychologists who work for the advertising industry, and even the education industry, and so forth, who have reason to study certain average behavioral dispositions of people quite apart from reference to the teleo-functional aspects of these behaviors. Also, but less happily, that there have been animal

studies done under the flag of behaviorism that involved extreme deprivations and other abuses to experimental animals with no thought given to whether the results obtained flowed from normally functioning mechanisms or, instead, from grievous damage to the animals' insides. Similarly, the Nazis are supposed to have used Jewish prisoners to study aspects of "physiology" with no concern about whether the effects they were observing were the result of mutilating the physiological systems or whether they were effects of normal, that is, adaptive, functioning under stress. But to study organisms in that sort of way, even for praiseworthy purposes such as the promotion of effective and efficient education, or the fostering of effective psychotherapy, is surely not the core job of the biological sciences. For example, physiologists and ethologists are usually concerned to study healthy animals rather than diseased or mutilated ones, and not because the healthy animal is the average animal. The point is not, in general, a quest for laws holding on the statistical average. Rather, the healthy animal is, indeed is by definition, the animal that is so constructed that its parts can perform each of their teleo-functions adequately, if given appropriate environmental contingencies. The healthy animal is the animal that does or could function normally in the normative sense of "normal."

The biopsychologist's study has little to do with averages over chunks of living matter. The subject of the biopsychologist's study is the stages of an ongoing cyclical historical process, an ongoing event in history. As such, it is not, strictly speaking, even the study of a secondary substance or of a natural kind. The chunk of matter, the exemplar of a natural kind, that is the current specimen on the table represents, if it is lucky, a partial cross section of the target event. It represents a stage in the historical cyclical process. It is an embryo stage, an infant stage, an immature or a mature stage, on its way to the ensuing stage.

Being more precise, it represents cross sections from a numerous set of loosely co-ordinated intertwined parallel processes, each having strands of its own, each developing through its own inner logic and at its own pace in rough harmony and interchange with the others. In the case of human cognition, for example, the various strands of the processes of perceptual learning, concept formation, the development of beliefs and desires, and of progressively more effective use of beliefs and desires through action, are roughly integrated with stages in the development of various motor skills, with many aspects of physical growth and development, and so forth. It is inevitable that certain strands of these processes should fail in the case of individual animals, and if failure is central and massive enough and redundancy in the system not sufficient to overcome it, the individual dies. The historical species, and to a lesser degree each individual animal, is like a rope with a small central core of overlapping strands running from one end to the other, the majority of strands, indeed the vast majority, being peripheral and very short where they have broken off. The biopsychologist's study concerns only the central unbroken strands of this fabric, and each fiber in these only so far as it has spun itself out in a principled historically precedented way. Such a study is not a study of substances or kinds, and not a study of averages.

V. The Organismic System Penetrates into the Environment

Because psychology is the study of processes resulting in and through external behavior, it focuses where the organism and the environment interlock, or better, merge. For there is no clear line but only the most arbitrary demarcation between the organism considered as a process and its environment. The organismic process has no skin. It is constantly sucking in matter from its surroundings and spewing it out again. Every breath is a refusal of separation from the environment. Nor are those aspects of the biological process that are cyclical in the sense of being reproduced confined within the skin. Spider webs and moth cocoons, bird nests and beaver dams, are reproduced by the genes out of environmental materials exactly as are bones, wings, and eyes. Richard Dawkins (1983) discusses the phenomenon of "the extended phenotype" through which boundaries between biological individuals or species become blurred, the biological projects of (the genes of) one individual or species being carried out through opportunistic manipulation of the bodies or behaviors of others. And he

discusses also the more obvious way in which phenotypes are extended into the environment through incorporation into the organismic system of inanimate non-body parts such as animal artifacts, and of other adaptive effects of an animal's behavior. The extended phenotype may thus reach yards or even miles beyond the animal's body. The unity of the organismic process might better be compared, then, to that of a wave or, say, a whirlpool, than to that of an ordinary physical object. Yet it is not as close, even, as a whirlpool to being encompassed within a unit space-time worm. For example, the beaver and his dam are aspects of the same organismic system, yet they are separable both in space and in time. And the beaver's dam is also part of the systems that comprise each of his kin.

Through its behavior, the biological system that is an animal merges into and incorporates portions of its environment. Inner mechanisms initiate processes completed by outer mechanisms, through outer structures and conditions that are either given in the environment or that have been put in place through prior behaviors of the individual or his kin. It is not just that the teleo-functions of an animal's behavioral systems are, as such, "long armed" functions. The animal itself, considered as a system of events, extends far out into the extra-body environment. To study an animal's behavioral systems without at the same time studying the normal integration of these into the environment, without studying the loops through the environment it is the function of these systems to initiate, would be exactly like studying the digestive system without considering what normally passes through it. Digestion without food is an exact analogue of behavior without environment. Turning to another analogue, to study behavior abstracted from the environment would be no less absurd, and for exactly the same reason, as if one were to study the structure and function of the heart's ventricles while ignoring the existence of the auricles and of the blood that passes through. The other half of the system containing the behavior-producing mechanisms lies in the environment in exactly the same sense that the other half of the system containing the ventricles lies in the atrium and in the blood running through.

Imagine attempting to study the inner mechanisms that produce migration in birds, or nest-building, or mating displays and female reactions to them, or imprinting in birds, without making reference to the way these mechanisms have historically meshed with the birds' environments so as to perform the functions for which they are named. More vivid, imagine attempting to study the origins of the co-ordinated motions made by the eyes and the head and the hand that effect eye-hand co-ordination while leaving completely out of account that there is, normally, a seen and felt object in the environment that mediates this co-ordination. It is equally ludicrous to suggest studying the deeper systems that produce human behaviors, for example, the systems that process beliefs and desires and intentions, etc., without considering how the environment has historically mediated performance of their functions. To understand what the ventricles do one must understand also what the auricles and the blood do and understand the relation that the ventricles ideally bear to the blood and the auricles. To understand what beliefs and desires do one must understand what the environment is doing and what relations beliefs and desires bear both to one another and to the environment.

It is always possible to describe any motion that an organism makes categorially. It is possible, anyway, to describe it relative only to the organism itself rather than relative to environmental structures.[7] And to describe sufficient causes of an organism's categorially described motions, one can always begin, merely, with categorial descriptions of the organism itself plus categorial descriptions of the environmental input to the organism. From this perspective, the organism's relation to its distal environment appears causally impotent in the production of its motions. But its motions are not its behaviors. The changes in categorial structure thus traced are significant biologically, are aspects of behaviors, only insofar as

[7] For a discussion of biologically relevant vs. biologically irrelevant ways of describing the same behaviors, see (Millikan forthcoming).

changing certain categorial properties of the organism effects significant changes in the organism's relation to its environment. Good comparisons are not easy to find here but we can try this one. The dentist has no particular interest in the exact shapes and sizes of individual persons' lower jaws. That is, he has no particular interest in them disregarding their relations to the shapes and sizes of the teeth that fit into them and of the upper jaws they must match. Clearly the details of absolute structure are not significant here, but only the details of relation. The same must surely be true both for the physiological structures supporting cognitive functions and for the behaviors these help to produce. Their relations to one another and their relations to the environment are what is biologically significant, and what is, for the normal case, uniformly describable within biological theory, not their categorial properties.

From the perspective of biopsychology it should be evident both that the personal history of the organism is relevant to understanding its psychological nature and that its way of interlocking with the environment is relevant. Whether a person remembers or merely seems to remember, whether a person knows or merely believes truly, are matters of personal history, but equally are matters of whether the biological cycle is proceeding normally or whether some of its links have been forged only by luck.[8] Whether a belief corresponds to the outer world as it should or is false instead is a matter of the interlocking of the organism with its environment, but equally a matter of whether the wider organismic system, which system includes part of that environment, is normally, that is, ideally, constituted or whether it is biologically abnormal. If it is biologically abnormal, this shows, in turn, that abnormalities must have occurred in the development of the system, for in so far as the organism-environment system cycled entirely normally (which, of course, it never does) beliefs would all come out true.

VI. Biopsychology is a Predictive Science if at all, then only Accidentally

These reflections on the nature of intentional psychology entail that, as a biological science, it does not aspire to be predictive. Biopsychology studies what happens when biological processes proceed normally, but the normal is neither the necessary nor always the statistically average. Prediction and control do of course play an important role under parts of the wide umbrella called psychology --I have mentioned psychological testing, human engineering, psychotherapy, etc.--but prediction and control are not required biproducts of intentional psychology. Indeed, intentional attitude psychology is a rather unlikely candidate to aspire to the detailed prediction of individual human behaviors.

This is true for at least two reasons. The first is diversity among individual constitutions. For psychology to predict individual behaviors, just as a starter, babies would have to be born cognitively and affectively, indeed also physically, alike. But it is abundantly clear that different newborns, inserted into identical environments, would not behave at all alike, unless under the most general and vacuous of descriptions. People are born with predispositions to different cognitive and affective styles, with different cognitive strengths and weaknesses. Non-psychological factors such as body-build, reaction time, energy level and health also play a large role in determining behavior. Further, it is likely that many aspects of our cognitive processes are partly stochastic, hence that which among many possible solutions to a given problem an individual discovers and executes often is not governed by well-defined psychological principles at all. Surely nothing short of complete physical and chemical analysis could in fact predict the detailed behaviors of any individual. The individual is not a replica of its ancestors or of its friends. It is a bundle of heavily redundant unfolding sub-systems adapted each to the others' concrete peculiarities to form co-ordinated larger units, this in accordance with principles of co-ordination and development all of which are, as yet, subject to merest speculation. We are still trying to find out how an individual's

[8] For a compatible discussion of the nature of knowledge, see (Millikan 1986b).

muscles and tendons grow the right length to fit the individual's bones, let alone how the various facets of individual cognitive development and function grow into a coherent unit. But there is no reason to suppose that exactly how an individual thinks is any more governed by laws quantifying over individuals than, say, how he walks or plays tennis-- or how he reacts to allergens.

The second reason that intentional psychology cannot be required to predict individual behaviors is that there is no compelling reason to suppose that all or even most of the norms that it describes are usually fulfilled. Most obvious, as has already been noted, is that the environment cannot always be relied upon to do its part in completing the functions of the cognitive systems. Because this is so obvious, it has been equally apparent to all that there could not possibly be any reliable laws of organism distal environment interaction, certainly not for the case of humans. Hence theorists who take it that psychology's main business is to deal in laws have found it necessary to insist that a scientific intentional psychology would have to be "narrow", that it would have to ignore the environment. But it is also likely that those portions of the cognitive functions that are carried out inside the organism are abundantly vulnerable to failure. The cognitive mechanisms seem to be paradigms of functional redundancy and layered back-up systems, commanding a variety of means to the accomplishment of the same or functionally equivalent projects. If at first you don't succeed, try another way, is a fundamental heuristic for our cognitive functions. Witness, for example, the well-documented variety of forms of compensation employed by those with brain damage. This redundancy strongly suggests the vulnerability of various cognitive techniques taken separately. It follows that there is little reason to suppose that the exact progression of anyone's inner cognitive systems could be predicted on the basis of even the most exact understanding of all types of human cognitive teleo-function, an understanding of all the biological norms involved.

Suppose, for example, that man is indeed a rational animal; that conformity to certain logical principles is a biological norm for human thought-processes. It would not strictly follow that conformity to reason was so much as a common occurrence. Reasoning could be one among other functions of the behavior-controlling systems, one that sometimes worked and was then to the organism's advantage. It could also be one that seldom caused irreparable damage when it failed, due to redundancy and to backing by various cruder behavior-controlling devices such as those found in the lesser animals. Indeed, remembering the way evolution works, it seems that there must at least have been a time when human reason had exactly this tentative status. It is not likely that the ability to reason well or to learn to reason well arrived all at once in a single lucky mutation. And we can raise the question of how well, in fact, the average modern human reasons. Clearly from the fact that drawing rational inferences may be a norm for the human cognitive systems it does not follow that any reliable predictions about inference patterns can be made. Even though man is a rational animal, rational psychology could remain very far from a reliable predictive science.

But a strong contemporary tradition has it that rational psychology must be a predictive science if it is to be a science at all, and that its central job is exactly to predict individual humans' behaviors. It is claimed, further, that our layman's way of thinking about intentional mental states constitutes a "folk psychological theory," the central employment of which is to effect prediction of the behaviors of our fellows, for this is necessary in order to project our own paths through the tangle of other folks' actions. Do I maintain that it is mere illusion that we thus predict the actions of others?

No, it certainly is not an illusion that we do a lot of correct predicting concerning the behaviors of others. Most ordinary forms of social intercourse and social co-operation would be impossible if we could not. But there may be a misunderstanding over the methods that we typically employ for prediction. The tool that we most commonly use, I suggest, is not a theory of the inner mechanisms that lie behind predicted behaviors. The tool is not, for example, belief-desire theory. Most of our predictions are done with a much blunter tool--the

method of brute correlation. In many cases there is, of course, some understanding of the outlines of the psychological mechanisms lying behind predicted behaviors, but our predictions do not usually rest on this understanding, either at all, or at least very deeply. They rest mainly on observations of past behavioral regularities for the individual and for the group(s) to which the individual belongs. Within fairly well-defined limits, people, especially people from the same culture, just do behave uniformly in a theater, on the road, at the grocer's, even when recreating in the park. Most people are more likely than not to meet what others consider to be their business and social obligations, to conform to general expectations concerning what is appropriate or seemly and, very important, to do the things they have said they will do. Beyond this, we project ahead patterns observed in the past for particular individuals. Known personality traits, character traits, and habits serve as our guides. Of course such knowledge merely limits the boundaries of people's likely behavior. It does little or nothing towards actually determining behavior in its variety. But seldom do we make an attempt to predict others' behaviors in much more detail than this. How inept we actually are at predicting behaviors, even of our best friends and family members, when these behaviors are not covered by known regularities, may be illustrated by friends who become separated in a large crowd, say at a fair, each trying in vain to outguess what the other will do in an attempt at reunion.

VII. Reasons and Causes

If we are rational, what that means is that rationality is a biological norm for humans, not that rationality is necessitated by special causal laws of human psychology. Compatibly, it is standard nowadays to claim (though on somewhat different grounds) that thoughts categorized in accordance with their semantics are not the sorts of things that could, even in principle, fall under causal laws. On the account of this essay, the semantic category of a thought is determined relative to its biological functions, which depend in turn upon its history, upon its place relative to certain prior events. But having a certain history is not, of course, an attribute that has "causal powers." Hence reasons can not be, as such, causes. More generally, that a thing has a teleo-function is a causally impotent fact about it. Especially, it is never directly because a thing has a certain function that it performs that function or any other function. More nearly the reverse is true. The thing exists and has a certain function because things homologous to it have performed that function (better, had that effect) in the past. Moreover, here the "because" is only partly causal, the other part is constitutive or logical.

But perhaps it will be thought that although things that have functions cannot be supposed to perform these functions either on account of having these functions or in accordance with strict causal laws, still they must perform them in accordance with ceteris paribus laws. Roughly, there have to be conditions under which the functional item would perform its functions since there have to have been conditions under which its ancestors did perform these functions, and the same kind of item in the same kind of conditions would do the same kind of thing again. This ignores defective members of function categories--diseased hearts, injured limbs etc. It also ignores the fact that performance of their functions is, for many items, a relatively rare occurrence. Would we really wish to speak of ceteris paribus laws in cases where ceteris non paribus sund most of the time? And it ignores also a third point.

Characteristically, the same function could, at least in principle, be performed by many differently constituted items. But if these items are differently constituted, if they operate in accordance with different principles, then the supporting conditions required for them to effect this function must differ as well. Brain cells performing the division algorithm require oxygen whereas computer chips require electric currents, and so forth. Similarly, the outer world conditions that support the bat's mosquito-locating abilities and those that support his mosquito-catching abilities are different from those that support the same abilities in humans. (The bat can perform in the dark on silent mosquitos; humans cannot.) The result is that there are no
ceteris paribus laws covering all items having a certain function. For ceteris paribus

conditions are unspecified conditions that must remain the same from case to case for the law to hold, whereas here the necessary conditions would have, precisely, to vary from case to case. A "law" applying to all such cases could say no more than that the items falling under the law could be made, by adding different circumstances tailored specifically to each case, to perform the function. But surely anything can be made to effect anything if one adds the right intervening media, if one adds enough special enough circumstances. So any such "law" would be empty. There are no causal laws of any kind, then, that directly concern the causal efficacy of reasons as such. The closest we could get would be ceteris paribus laws for human reasons, other ceteris paribus laws for dolphin reasons, still others for Martian reasons, and so forth.

VIII. Normalizing Explanations

Our argument suggests that explanation of an agent's behavior by reference to reasons for acting is not best analyzed as explanation by subsumption under causal laws. The question that arises then is what kind of explanation the citing of reasons for acting is, and how it can still be causal-order explanation. Intentional attitude explanations of behaviors proceed, I will argue, by subsumption of behaviors under biological norms rather than laws, and/or by noting departures from these norms and, perhaps, causes of these departures. Following Philip Pettit (1986), to whose views mine run parallel here, I call such explanations "normalizing explanations." The status of explanations of individual behaviors by reference to reasons concerns the relation of normalizing explanations to other forms of causal-order explanation that are, perhaps, better understood.

To explain a phenomenon by subsuming it under norms is to exhibit it as an instance of conformity to or departure from proper operation of some teleological system. A very simple form of normalizing explanation explains the occurrence of a phenomenon by reference, merely, to something whose function it was to produce that phenomenon. For example, the dishes are clean because they have been put through the dishwasher; the washing machine door is locked because the washer is not finished spinning and the door is designed not to unlock until it is finished spinning; the bear is asleep because it is winter and it is (biologically) normal for bears to sleep through the winter (see note 4 above).

In order to explain a phenomenon this way it is necessary, of course, to classify it appropriately as the outcome of a teleo-functional process, and this classifying may itself count as a simple form of explanation. What is happening? What is it doing? It's washing dishes, not making soup or just dirtying the water; it's winding a magnetic coil, not storing wire on a spool; it's resting, cooling its motor between cycles, not playing dead, or broken; and so forth.

More complex normalizing explanations tell or implicitly refer to the place an event has in a series or interdependent pattern of functions, or tell where, and perhaps why, malfunction occurred within such a series or pattern. Thus, that cog-wheel's turning in the calculator is its carrying one in a certain addition algorithm; the car went through the light because its brakes failed; the outboard stalls because there's dirt in the carburetor that gets into the needle-valve.

Normalizing explanations often make reference to conditions that must be presupposed for normal operation of a device or system. Thus the outboard won't start because the spark- plugs are wet or because there's no gas it the tank, the scuba-diver passed out because it was too cold or because his tank ran out of oxygen, and so forth.

Finally, the relation between certain conditions of the functional system itself or of the environment and certain states of the system that normally adapt the system's progress to those conditions may be targeted in a normalizing explanation. Thus, the motor is racing because the heavy-load switch is on but the load is not heavy; the washer failed to fill properly because the soap was put into the tub rather than into the dispenser so that the rising suds tripped the water cut-off before the tub was full; the animal's winter-approaching detectors failed because it was kept indoors, which is why it is attired inappropriately or is behaving inappropriately to the season.

IX. The Normal Roles of Beliefs and Desires

Notice how natural it would be to say in the last two of these cases that the washer thought it was full when it wasn't and that the animal's system didn't know it was winter. This is because a belief or a bit of knowledge is likewise a teleo-functional item, one whose function is to adapt the containing system so that it can perform its functions under certain conditions, namely, those conditions which the belief is about. Or, being a little more precise, it is the belief-forming mechanisms that produce the adaptations, the adjustments of the organism to the environment, the beliefs. Beliefs themselves are functionally classified, are "individuated", not directly by function but according to the special conditions corresponding to them that must be met in the world if it is to be possible for them to contribute to proper functioning of the larger system in a historically normal way. Somewhat similarly, the water switch's being off will promote the washer's tasks normally only if the condition is fulfilled that the washer is full. And the animal's winter detectors' being off will effect appropriate functioning of the animal in accordance with historically normal reasons only if winter is not yet approaching.

Explicit human beliefs, however, are much more than just biological adapters to certain environmental conditions. They are adapters that perform their tasks in a certain sort of way, namely, through participation in inference processes. A picture that I advocate but will not try to defend here (see Millikan 1984,1986a,1989a,1991b) shows beliefs and desires as working for the organism by modeling (in accordance with very abstract mathematical mapping functions) the environment, modeling the organism's goals, and modeling types of environmental transitions that the organism knows how to bring about. Normal practical thinking then involves tinkering with these models until solutions are found that will effect transitions from the present state of the environment to various desired states. On this picture the teleo-functions of desires (which they may not very often perform), like those of blueprints, are to effect what they model, to get themselves realized. When everything goes according to norm, action guided by the models inside is action conformed to the outside world so as to issue in productive loops through the environment. This happens in accordance with explanations that, made fully general, that is fully spelled out relationally, apply perfectly generally to all successful uses of the (same capacities of the) species' cognitive systems, historical and current. Theoretical inference is then interpreted as a process whereby the internal model of the environment grows or extends itself in accordance with principles that model various logical, geometrical, and causal necessities or regularities or dependencies in the environment.

Be all this as it may, what seems quite certain is that there must exist some sort of systematic teleo-functional organization of the human cognitive systems whereby the making of good practical and theoretical inferences corresponds to normal (but perhaps not average) functioning for beliefs and desires, and whereby it is biologically normal (not average) for desires to be fulfilled, at least under certain conditions. (Why else the capacity to have desires?). Accordingly, explanations of behaviors by reference to reasons for action are normalizing explanations.

X. How Normalizing Explanations Circumscribe Causes

Why it is that normalizing explanations explain, how it is that they fall under a general theory of explanation, is too large a question for this essay. Our question here is only how such explanations connect with simpler kinds of causal explanation. One connecting link is that whatever has a teleo-function has a normal way of operating, a normal way of performing its function. For functional artifacts this may be, in part, the way the designer proposed that the function be performed, for biological devices, it is the way the function has been performed historically. An exhaustive analysis of the way, given its history, that any functional item operates when operating normally, arrives eventually at a description of normal physical structure for such a device and normal physical conditions for its operation, such that physical laws generate performance of this function given this structure and these

conditions. By making implicit reference to such causal explanations, normalizing explanations may thus circumscribe quite specific physical explanations without detailing them.

Guided by Cummins (1975, 1983), we notice that the analysis of how a system normally functions may have several parts. First, the larger function or functions of the system may be analyzed into sub-functions that are performed either serially or simultaneously or in some more complicated pattern of interaction. This kind of analysis Cummins calls "functional analysis." Cummins suggests that a functional analysis may generally be represented by a flow-chart, but of course highly parallel processes, especially those that interact to some degree stochastically, must be represented otherwise. Second, the system may be analyzed into sub-systems, which may or may not correspond to discrete physical parts, each of which is responsible for a designated set of sub-functions. This kind of analysis Cummins calls "compositional analysis." Compositional analysis results in a description of the normal (not necessarily actual) constitution of the system by reference to parts described teleo-functionally, that is, normatively rather than dispositionally. (Here I depart from Cummins, who equates functions with dispositions.) Finally, the normal physical constitutions of the elements normally composing the system may be described, along with the surrounding physical conditions required for normal functioning, and it may be shown how these descriptions together account, in accordance with physical law, for cases of normal operation. That is, the system may ultimately be analyzed into a set of physical parts and physical dispositions rather than, merely, functionally categorized parts and normal functions.[9]

By reference to the possibility of this kind of physical analysis, explanations of behaviors according to reasons for action may circumscribe physical causes. Compare explaining why a man shakes by saying that he suffers from Brown's syndrome, even though the etiology of Brown's syndrome may not be known. Or compare explaining why a man has brown hair by saying he has genes for brown hair rather than, say, having dyed his hair, though no one knows the constitution of the gene or how it produces brown hair.[10]

That this is not the complete answer to how reasons circumscribe causes becomes evident, however, when we remember that devices falling in the same function category can have widely varying constitutions. For example, we do sometimes explain, say, how John managed to get the can open by noting that he finally found a can-opener, but given the enormous variety among can-openers, the various different principles on which they may work, how could such an explanation possibly do anything towards circumscribing physical causes or types of physical processes lying behind the can's having come open? Similarly, if there really were various other creatures designed quite differently from humans and made of quite different stuffs but who still had beliefs and desires, then explanation of actions by reference to beliefs and desires without mention of the species of creature involved would seem not to circumscribe any particular kind of physical process at all (cf. Block 1990).

But, looking more closely, whether it circumscribes a kind of process depends on how you

---

[9] This does not imply that, given a certain species, there is a classically understood type-type identity relation between, say, normally constituted and normally functioning beliefs and desires about x on the one hand, and certain physiological structures on the other. Certainly if the physical constitutions of human beliefs are typed categorially there is no reason at all to suppose that any such identity holds. If there are bridge laws for humans that map the semantics of thoughts onto physiological structures, surely what these laws map is certain semantic relations among beliefs and desires onto physical relations among these, hence principles of logical interaction onto principles of causal interaction, not categorial meanings onto categorial physiological "shapes."

[10] Compare (Block, 1990).

type your kinds. Behind every normalizing explanation is a device or system with teleo-functions, and an item acquires a teleo-function only by having a very special sort of causal history. For example, if the cat's purr is explained as produced by a purr-box, an organ especially designed, in the smaller cats, to produce purrs, then we know that the purr-box itself has resulted, ultimately, from the operation of prior purr-boxes in ancestor cats which produced purrs, these purrs somehow having survival value, contributing an essential link, at least occasionally, to the historic cat-chain. Thus a salient cause of the purr is a series of prior purrs. Of course when the functions referred to by normalizing explanations are described categorially though they are actually derived from relational functions, no such simple analysis applies. Still, to assign to any phenomenon a place in a functional system is to claim that it has emerged from a very special kind of causal-historical process, a kind that defines functionality. It is to distinguish its particular type of causal origin quite sharply from other etiological patterns.

BIBLIOGRAPHY

Block, N. 1990 Can the Mind Change the World? in G. Boolos, ed., Meaning and Method; Essays in Honor of Hilary Putnam (Cambridge University Press).

Cummins, R. 1975 Functional Analysis Journal of Philosophy 72:

---------1983 The Nature of Psychological Explanation (Bradford Books/MIT Press).

Dawkins, R. 1983 The Extended Phenotype (Oxford University Press).

Dretske, F. 1988 Explaining Behavior (Bradford Books/MIT Press).

Gould, S. J. and Lewontin, R. C. 1979 The Spandrels of San Marco and the Panglossian Program. Proceedings of the Royal Society of London 205:281-288.

Jackson, F. and Pettit, P. 1988 Functionalism and Broad Content Mind 97 pp. 381-400.

Kim, J. 1989 The Myth of Nonreductive Materialism Proceedings of the American Philosophical Association 63.3:31-47.

Millikan 1984 Language, Thought and Other Biological Categories Bradford Books/MIT Press.

---------1986a Thoughts Without Laws: Cognitive Science with Content Philosophical Review, 95.1:47-80.

---------1986b Naturalist Reflections on Knowledge Pacific Philosophical Quarterly 65.4:315-334

---------1989a Biosemantics The Journal of Philosophy 86.6:281-297.

---------1989b In Defense of Proper Functions. Philosophy of Science 56.2:288-302.

---------1989c An Ambiguity in the Notion "Function" Philosophy and Biology 4.2:172-176.

---------1990a Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox The Philosophical Review 99.3:323-353.

---------1990b Seismograph Readings for 'Explaining Behavior', Philosophy

and Phenomenological Research 50.4:807-812
---------1991a Speaking up for Darwin, G Rey and Loewer, B. eds., Meaning and Mind, Fodor and His Critics, Blackwell
---------1991b Compare and Contrast Dretske, Fodor, and Millikan on Teleosemantics, Philosophical Topics
---------(forthcoming) What is Behavior? In White Queen Psychology and Other Essays for Alice, Bradford Books, MIT Press.
Pettit, P. 1986 Broad Minded Explanation and Psychology, in P. Pettit and J. McDowell, eds. Subject, Thought, and Context (Oxford, Clarendon Press):17-58.