MIDWEST STUDIES IN PHILOSOPHY, XIX (1994)

Voluntary Motion, Biological Computation, and Free Will

PATRICK SUPPES

FREE WILL

What bearing does the general indistinguishability of determinism and indeterminism, as I argued for in a recent paper (Suppes, 1993), have on the classical philosophical problem of free will? Different answers, it seems to me, are possible. A conservative response might be that it has no bearing or relevance. According to this conservative response, free will is a manifestation of the causal efficacy of mental events. Such events are, in turn, manifestations of physical events, which unfortunately cannot be observed directly. But this language, suggesting, as it does, two kinds of objects, the mental and the physical, will be objected to by some. There are, so it would be said, just physical events, but under some descriptions we speak of them as mental events. In any case, I set this issue concerning the kinds of events to the side. This meant-to-be straightforward causal account of free will leaves no place for indeterminism or instability in the arena of human intention and action, even though it is entirely compatible with indeterminism in the behavior of elementary particles or black holes.

In contrast, the radical response is that the exercise of free will, as seen in the daily activity of humans and other animals high on the evolutionary scale, constitutes prima facie evidence of indeterminism in human affairs. This radical view, now not much defended by many philosophers, would obviously favor indeterminism over instability as the correct philosophical backdrop for the plain man's view of the world.

I do not intend these rough sketches of contrasting responses to be taken seriously. Rather, my purpose is to try to make clear a position that does not commit to either determinism or indeterminism in its account of free will. The attempt to be independent of either is just because of my earlier argument about the impossibility of providing data to choose between them. I argued in the

essay referred to above that the choice between determinism and indeterminism is, at a fundamental level, transcendental.

There are, it seems to me, two central principles that should govern our account of free will. The first is that small causes can produce large effects.[1] The second is that random phenomena are maximally complex, and it is complexity that is phenomenologically evident in many human actions that are not constrained but satisfy ordinary ideas of being free actions. There is a third methodological principle I want to state before saying something more about these two principles. That principle is not to be intimidated by the kind of mechanistic argument derived from classical mechanics. Such arguments overwhelmed Kant in his metaphysical foundations of natural science and in his antinomy of freedom of the will. The case for determinism is not a strong one. The case for predictability of most phenomena is overwhelmingly a bad one. We need not be pushed by some metaphysical view of physics into being forced to create a dichotomy between deterministic physical processes and free human actions that seem to violate physical causality.

With these remarks as starters, I can sketch at this point the argument I want to make about free will. The central idea is that human intentions and actions exhibit in some cases the stability we anticipate of many parts of the physical world and in other cases the instability that is characteristic of many other kinds of physical phenomena. In human actions that appear unstable, either in their underlying intentions or in their triggering movements, we cannot in any direct way distinguish whether they are deterministically caused but unstable or, on the other hand, indeterministic. I said this before but in this context it is a litany that needs repeating. What we can say is that there are classes of human actions that are highly complex and essentially unpredictable. I give several examples in Suppes (1985).[2]

For the moment, I want to restrict myself to a single point, but an important one, that should illustrate my more general thesis about free will. In the case of movement of a limb, whether an arm or a leg, there are three kinds of actions or events that we can distinguish. One class consists of the movements themselves. Another consists of the forces exerted by the muscles, for example, a change in muscle tension to produce a movement, and the third consists of proprioceptive signals from the nervous system to the muscles to affect the forces exerted by the muscles. There is, of course, a great deal more to the system, at a gross level even, than I have indicated. For example, there is perceptual feedback generating the proprioceptive signals, but I shall ignore for the moment this additional critical perceptual part of the system. The important point is that there is no one-one correspondence between muscle force and movement. In fact, the evidence is very good that the differential equations are nonlinear in character, and have, as would be expected, extraordinarily complex solutions that are very dependent on initial conditions. The instability of the relation between forces and muscles requires constant revision of muscle tension by proprioceptive signals being sent to express the animal's intention in terms of what is desired as movement.

Now it might be thought that the aim of this kind of physiology and psychology is a mechanistic reduction of movement to an analysis of physical causes. But this is precisely what is not the case in any simple way. The extraordinary complexity of the system makes it impossible to pin down in any detail some simple mechanistic theory of the causal sequence. Whether the system is at heart deterministic but unstable or indeterministic cannot possibly be decided, and there is no hope with the framework of present science that the issue could be decided in decisive detail. The complexity is so great that there is little doubt that a strong argument could be made for irreducibly random elements in the phenomena, even though the central tendency of the mover's intention seems plain enough. One way of putting the indeterministic thesis in the present framework is that at heart the system is highly indeterministic but by continual proprioceptive correction the movement is kept on course.

It is this last point that I want to analyze with some care, for I think it contains the basis for one of the strongest arguments for the existence of free will regardless of whether "surrounding" phenomena are deterministic or not.

For much human and other animal behavior, continued perceptual feedback and proprioceptive correction of movement are obviously required. The first-level explanation of the ubiquitous perception-and-movement-correction data is in terms of goals and intentions. "Why is Mary turning the corner? Because she is going to the store." "Why is Henry moving backward so fast? He is trying to get to John's lob in time to hit the ball back across the net." This intentional feature of commonsense psychology is really denied by no one.

Such ordinary intentional actions bear the strongest possible witness to the existence of free will in the present framework. It is not the intentional aspect in its pure form that makes the case for free will. Rather, it is the ubiquitous necessity of correcting any complex movement continually in order to achieve the intended goal. Willful intervention, so to speak, is needed from moment to moment. The fact of such intervention is evident. Just as important, neither determinism nor indeterminism as a theoretical framework can shake the need for intentional intervention.

There is one fantasy about determinism that must be laid to rest. What is deterministic, so the fantasy goes, is not just the body that is moving, but the much larger system of the body plus environment plus perception plus feedback. This complicated total system is supposed to go its merry way without purposeful intervention. From what we know of complex systems, nothing could be less likely. Certainly we shall never, within the framework of present science, have the slightest chance of understanding in noninterventional or nonintentional terms even the simplest cases of human movement. But, given a goal, the phenomena become manageable even if still subtle and complex in physiological detail.

To believe in determinism simpliciter is on a par with believing in an omniscient God. In the real world of natural, as opposed to supernatural, phenomena, intention and will are evident and unproblematic. Intentional action is indispensable whether the world is indeterministic or deterministic. Which it is very likely we shall never know.

SOME QUESTIONS ABOUT VOLUNTARY MOTION

I now want to turn to a more systematic analysis of the general ideas just presented. Although the issue of free will or purposeful action can be analyzed from the standpoint of many different kinds of phenomena, ranging from the physiology of the heart to the evolution of the species, I want to use my earlier analysis of determinism and prediction (Suppes, 1993), and to concentrate just on the case of physical movement. There is, of course, already a large philosophical literature discussing human action involving physical movement, but it will not be my purpose to give a formal characterization of action but rather to concentrate on the more restricted notion of purposeful or voluntary movement.

The intuitive idea is quite simple. I shall define a physical movement of an object as *voluntary* if the movement of that object cannot be predicted by knowledge of initial and boundary conditions and laws of physics, but can be predicted by knowledge of a goal that is intrinsic to the object or that is set for this object by some other agent.

There are a number of distinctions that need to be dealt with, and unfortunately not all of these distinctions can be handled in an uncontroversial manner. I mean by this that serious arguments can be found on both sides as well as serious proponents supporting these arguments, and it is not clear that every issue can be resolved in an objective fashion. However, in order to make the case for the existence of free will, it will not be necessary to have a settled view on each of these matters. Let me mention some of the principal issues as I see them.

1. Is the motion of low-level organisms such as parameciums, which can be conditioned, to count as voluntary?
2. Is the motion intrinsic to a homeostatic device, such as a household thermostat, to count as an example of purposeful motion?
3. Is some at least rudimentary form of perception required to provide feedback and correction of the course of motion necessary in order for a motion to count as purposeful or voluntary?
4. Must an object exhibiting voluntary behavior be described as having intentions?
5. Does voluntary motion of an object imply consciousness of a goal by the object or an agent controlling the goal?
6. What level of accuracy of prediction is considered acceptable for purposeful or voluntary motion?

I shall consider these six issues in reverse order.

Accuracy of prediction. If my wife or I go to the nearest grocery store to purchase a bottle of milk and a loaf of bread, I can predict within a few feet or meters where the motion will stop, and then return to the house, again to a position known within a few meters or feet. I can, of course, even pick a few symbolic points, for example, the door of the grocery store or one of the doors into the house, to fix the accuracy of prediction well within a meter. Such

predictions are pretty gross by the standards of good physics or engineering. If I ask for a more detailed prediction of the space-time trajectory—which would be the ordinary object of physical prediction—to the store and back, then quite gross errors can occur. First of all, the system is an open one, in the sense that there are disturbing causes easily entering the simple system defined. A typical example would be a traffic jam that had not been foreseen. Even the timing of approach to the automatic stop lights en route would lead to considerable variability in prediction. What is most important, however, is the intrinsic variability. There is no possibility, from physical knowledge of the initial conditions, as I or my wife leave the house, to calculate with any accuracy whatsoever the space-time trajectory to the store and back. The actual trajectory followed will require continual correction and modification based upon perceptual feedback of a variety of kinds en route. Thus we may conclude that prediction using just laws of physics is absolutely out of the question. There is no hope, using just the laws of physics and initial conditions, of being able to predict when someone would arrive at the entrance to the store, at a point whose location is known, within much less than a meter. On the other hand, the introduction of purpose and the concept of a goal do not make prediction come close to that ordinarily to be found in good physics or engineering, as already mentioned. What we get, with the introduction of purpose, in the sense of physical behavior oriented toward a goal, is that we can make very good predictions about one or two special aspects of the motion, namely, the aspects most central to achieving the goal. We can also, in terms of implicit aspects of the goal, make some quite reasonable predictions, but not highly accurate ones, about such features as the time the round trip will take. As already remarked, the pluralistic and decentralized character of the human motor-control system also argues against any possibility of highly accurate predictions of trajectories of purposeful motion. The important point, however, is the strong contrast between the essential unpredictability by physics of the critical points of the motion and the relatively accurate prediction of these critical points once a goal is introduced.

A second fundamental point is that in the framework of present science we are utterly unable to develop a detailed and accurate theory of voluntary motion. Moreover, it seems in principle impossible, and a mistaken idea to attempt to realize. The failure, it is important to note, can already be found in the analysis of complex open physical systems without any component of voluntary motion.

The voluntary aspect makes it all the more out of the question to develop a detailed trajectory prediction. From a physical standpoint, the course corrections characteristic of voluntary motion increase the openness of the system, and thereby increase the problems of physical prediction.

Consciousness. As I define the notion of voluntary motion, the requirement of consciousness is not part of the definiens. I have in mind examples that I want to include in which consciousness is certainly not an unequivocally definite aspect of behavior. When a bird dives from a hundred feet into the ocean to catch a fish it has seen, its complicated motion is certainly purposeful

in both the weak and strong senses that I would want to use, but it is doubtful if we want to attribute consciousness to birds. It is not even clear that we want to attribute consciousness to dogs. A dog that is told to fetch the paper, certainly in the sense used here, engages in purposeful motion, but may not be conscious of the goal he is achieving because there is not the appropriate concept of consciousness to apply to the dog's mental life.

Intentions. Obviously I do not intend to make conscious intentions part of purposeful motion. On the other hand, a derivative sense of intention implicit in all goal-seeking behavior is innocent and unexceptionable, but the important point is that I do not intend to introduce a separate concept of intention that does any serious work in what is to follow.

Perception. Perception is another matter. Without perception and the consequent feedback mechanisms adjusting the motion of objects seeking a goal, I see no way to clearly distinguish purposeful from nonpurposeful motion, and therefore to distinguish physical predictability from what I shall term purposeful predictability. The way in which I or my wife depend upon perception to negotiate the physically complicated path from our house to the grocery store is obvious and certainly not a matter of essential controversy. On the other hand, as we move down the phylogenetic scale and on out to simple homeostatic mechanisms there can be differences about what is to be regarded as perception. In the case of lower animals, we can certainly go a long way down the phylogenetic scale and still be forced to accept that some sort of perception is taking place, because specific sensory input is demonstrably affecting the motion the animal is engaged in. The perception, and the perceptual system supporting these activities of the organism, can be extraordinarily simple in comparison with our own rich perceptual apparatus, but still I would claim it is perception when an ant follows a trail in pursuit of food or in returning to the nest. The acute visual perception of birds is well known.

Thermostat. But what about the thermostat as a simple example of an homeostatic system? It certainly is given a clear goal by the agent controlling it, namely, keep the ambient air temperature around the location of the thermostat itself in a certain small interval of values. Second, the thermostat has a rudimentary form of perception, for it is able to 'perceive' the temperature of the ambient air surrounding it much more accurately than a person can. It also can be claimed that it is not a simple matter whatsoever to predict from initial and boundary conditions at time t_0 when the thermostat will activate the furnace (or the air conditioner), or, if the furnace is on at t_0, when it will turn it off. In that sense the motion of the thermostat, that is, the mechanical motion of its parts, including the turning on of the furnace, would seem to qualify as voluntary motion—I leave aside the fact that much of the physical activity of the thermostat is electronic rather than mechanical, a point that does not bear directly on the issue at hand. Perhaps the disturbing point about the thermostat is that it is completely controlled in terms of its mode of operation by another agent who has the freedom to set the constant value the thermostat is striving to

achieve, that is, its goal. We also find disturbing the fact that the design of the thermostat seems so transparent and understandable in purely physical terms. It is this absence of transparency, in fact, the absence of any serious detailed analysis at all, that is certainly responsible for much of what we consider important and significant about voluntary movement in higher organisms. It does not seem to me to be important to quarrel about exactly where to draw the line. I am prepared to say that in a weak sense a thermostat exhibits voluntary motion. On the other hand, I do want to stress that the thermostat is not an ideal example of a very simple object engaging in voluntary motion, because the thermostat itself is not moving but only its parts. I really want to restrict, for purposes of simplifying and making an analysis of one aspect of free will, the case of voluntary motion, where what one has in mind is the actual physical motion of the object as a whole. I have discussed the thermostat at some length just because it is a classical stalking horse for the sorts of philosophers who believe that intentions and free will are only embodied in complex structures of amino acids and their surrounding cellular materials.

MORE REMARKS ON VOLUNTARY MOTION

A methodological point. In offering an analysis of voluntary motion I am certainly operating in the philosophical rather than the scientific tradition. The systematic unsatisfactory character of my definition from a scientific or formal standpoint is evident. Unfortunately, the labor of converting it into something fully satisfactory is far from trivial, at least as far as I can see. On the other hand, the key notions used in the earlier informal definition are easy to describe, namely,

(i) the concept of predicting by standard physical laws and initial physical conditions,
(ii) the concept of having a goal, and
(iii) the concept of adjusting motion on the basis of perceptual feedback.

I have made the conditions only sufficient and not necessary, just to commit myself to less. Concerning (i), in the present case what one has in mind is relatively quite simple. One neither knows nor can discover by physical means a goal that a moving object is attempting to achieve either through its own volition or through that of an agent controlling it. The problem is expressing that idea in some reasonable way. I have left it vague and general as stated.

The prediction requirement in (i) may seem relatively definite because of the reference to "standard physical laws," but in fact such a reference is clearly quite nebulous and indefinite and any satisfactorily detailed theory, about which we might prove some theorems, would require a specific narrowing down of the physical laws or theories available.

Clause (iii) uses the intuitively obvious but formally vague notion of perceptual feedback and the equally intuitive but equally formally vague notion of adjustment of the trajectory of the object. Again, it is clear to all what one

has in mind, as in the case of an individual walking, running, or swimming to a given location that is a goal.

The analysis of voluntary motion was already a major topic in ancient philosophy, for it received systematic attention from Aristotle, and others as well. It is not in order here to look at the history of the concept, except to note that it is very old, and I do not mean to suggest for a moment that the idea of voluntary motion is a new one. The critical new ingredient I have introduced is the concept of unpredictability from the standpoint of standard physical theory, but even then there are clear harbingers of it in past analyses of voluntary motion. Of course, Aristotle does not discuss predictability of one kind or another, but rather the different kinds of causes that separate voluntary motion from involuntary motion. It is just part of my separation of determinism and predictability to put the emphasis here on predictability. In disentangling the issues concerning free will and determinism, in my own judgment the question of predictability is the most central.

Biology vs. ethics. Much of the philosophical literature on free will has been concerned with the relation between moral responsibility and freedom of the will. This is not merely a recent trend of modern moral philosophy but is already very much in the center of Kant's elaborate analysis of how freedom in the noumenal world can be seen as consistent with causal determinism in the phenomenal world.

In my view the emphasis on moral questions has been a disaster for the analysis of the problem of free will. Moral philosophy remains the soft underbelly of philosophy. The concepts and arguments of ethical thought are no match for the elaborate intellectual constructions of classical and modern physics, and their developed foundations. This intellectual mismatch may be seen already in Kant's placing on the side of experience all of physics, but essentially nothing of traditional philosophical concepts of moral behavior.

Aristotle had a more realistic approach. The expression of will—to use the modern term—is to be found first of all and most generally in the biological behavior of animals, not simply in the highly restricted domain of moral choices. Voluntary motion in animals is guided by reason, desire, or imagination. More generally, the analysis of biological structure and function is dominated by teleological principles.

As should be evident, I think that biological behavior of animals of all or nearly all species is the locus classicus of voluntary behavior. The analysis of voluntary motion alone shows the unfeasibility of reducing biology to physics. The next section will examine in more detail how this viewpoint recasts the philosophical problem of determinism and freedom of the will.

DETERMINISM AND FREE WILL

I now consider in more detail the relation between determinism, prediction, and free will, especially as exemplified in voluntary motion. There are several different ways to begin, but for both historical and conceptual reasons it will be useful to start with Kant's discussion in the Transcendental Dialectic of

the *Critique of Pure Reason*, especially in the section on the explanation of the cosmological idea of freedom in connection with the general necessity of nature. As might be expected, I want to oppose Kant's conclusions, but not always his methods. First, let us consider in somewhat more detail what he has to say. I assume that we are all more or less familiar with the general idea that the empirical world of phenomena, covered by the laws of physics essentially, is quite distinct from the timeless world of noumena in which we find the proper setting for reason and freedom. Here is what Kant has to say in one important passage for our purposes:

> Every man therefore has an empirical character of his (arbitrary) will, which is nothing but a certain causality of his reason, exhibiting in its phenomenal actions and effects a rule, according to which one may infer the motives of reason and its actions, both in kind and in degree, and judge of the subjective principles of his will. As that empirical character itself must be derived from phenomena, as an effect, and from their rule which is supplied by experience, all the acts of a man, so far as they are phenomena, are determined from his empirical character and from the other concomitant causes, according to the order of nature; and if we could investigate all the manifestations of his will to the very bottom, there would be not a single human action which we could not *predict* with certainty and recognize from its preceding conditions as necessary. There is no freedom therefore with reference to this empirical character, and yet it is only with reference to it that we can consider man, when we are merely *observing*, and, as is the case in anthropology, trying to investigate the motive causes of his actions physiologically. (A549–550)

In this passage I have emphasized *predict* although Kant does not. This passage generally exemplifies what I just said in brief about Kant's ideas, namely, the separation of the phenomenal from the noumenal world and the fact that the empirical character of action must lie in the phenomenal world. The point I want to especially focus on here is Kant's assurance that we could predict with certainty any single human action if we investigated all the manifestations of the individual's will, as he says, "to the very bottom." What he means here, of course, is under the empirical character of the action, which would mean an investigation according to the laws of nature governed ultimately by physics. It is exactly my thesis that this conception of Kant rests on a mistake. We cannot, in fact, predict human action with certainty, even under the most idealized circumstances. Above all, we cannot analyze an action so that it becomes determined necessarily by preceding physical conditions.

At the end of the part on freedom from which I am quoting, Kant has this also to say:

> Our problem was, whether freedom is contradictory to natural necessity in one and the same action: and this we have sufficiently answered by showing that freedom may have relation to a very different kind of

> conditions from those of nature, so that the law of the latter does not affect the former, and both may exist independent of, and undisturbed by, each other.
>
> It should be clearly understood that, in what we have said, we had no intention of establishing the *reality* of freedom, as one of the faculties which contain the cause of the phenomenal appearances in our world of sense. For not only would this have been no transcendental consideration at all, which is concerned with concepts only, but it could never have succeeded, because from experience we can never infer anything but what must be represented in thought according to the laws of experience. (A557–558)

In this passage especially, Kant makes very clear his insistence that freedom lies outside the necessity of natural laws, and, more generally, outside of experience.

It has been a familiar thesis of a large number of philosophers that determinism is incompatible with free will, which is what Kant is saying at the level of phenomena, and that therefore so much the worse for free will, because it is well established from physics that the world is deterministic. As Jules Vuillemin has remarked to me on various occasions, the central argument that I want to make is to stand Kant's view on its head. It is free will that is given as the central, empirical data, and it is determinism that is fragile, insubstantial, and unsupported, in any direct way.

Remember also that I am making the argument here not upon the more nebulous ideas sometimes associated with free will but with the necessity of voluntary motion exhibiting motion that is continually corrected as a purposeful matter in order to achieve a goal. So my claim is that, above all, the central empirical phenomenon that is undeniable is that of continual correction of voluntary motion based on perceptual feedback and proprioceptive control of locomotion, in order to achieve a goal, that is, a new location of the organism as a whole, or some significant part of the organism, such as a hand, leg, or eye. Furthermore, it is easy to find a wide class of voluntary motions because of their demonstrable unpredictability from the standpoint of pure physics.

Following through on Kant's ideas, let us say in turn what is to be said, on the viewpoint I am defending, about determinism. Here, determinism assumes the metaphysical will-o'-the-wisp status that Kant assigns to freedom. Determinism is compatible with empirical data of every conceivable kind, but it is in no sense scientifically established as a correct, quite general theory about the nature of the universe. Separation of prediction and determinism in the decisive way illustrated by many examples discussed in Suppes (1993) makes it appropriate to say that the assumption of determinism as a completely general thesis about nature is a transcendental metaphysical assumption that goes beyond experience. It is only through the possibility of prediction for phenomena of every significant natural kind that we can hope to sustain in a detailed and serious way a scientific thesis about determinism. Given the impossibility of such prediction, determinism must have a transcendental status.

Indeterminism. It is a familiar form of philosophical argument to show that free will is incompatible with determinism, and also free will is incompatible with indeterminism. Therefore, there can be no free will. But the kind of argument I have given about determinism I also make about indeterminism, which as a universal thesis has a similar transcendental character.

There is, of course, the question of how to interpret quantum mechanics and whether there is at rock bottom an indeterminism in quantum mechanics as exemplified by the work of the last two decades surrounding Bell's inequalities and Bell's theorem. Even here the argument is not at all decisive. For a recent spirited defense of the de Broglie–Bohm deterministic interpretation of quantum mechanics, see Holland (1993).

So the Kantian thesis has been properly stood on its head. Free will, as exemplified in voluntary motion, is the hard empirical fact. Determinism (or, if you prefer, indeterminism) is the transcendental metaphysical assumption out of reach of detailed confirmation.

BIOLOGICAL COMPUTATION

The positive theory of voluntary motion cannot depend only on negative results about physical unpredictability. The best short description of the new biological concept that is needed is that of computation. I want to insist on computation taking place in the behavior of even very simple organisms or agent-designed devices like thermostats. Perception, simple or complex, is, from the viewpoint I am adopting, a special case of computation. On the other hand, as should be clear from what I said earlier, consciousness or intention in any rich human sense is not a necessary condition of computation. Because of my concentration on voluntary motion I put to one side whether genetic transmission in the form of DNA code is a form of computation, although it is my own view that it is.

The three basic kinds of computation that occur in voluntary motion are:

(i) Setting a goal;
(ii) Perceiving something—object or phenomenon—in the environment;
(iii) Correcting the current trajectory.

Degenerate or special cases of this threefold analysis of computation in the context of voluntary motion are easy to give.

Stationary frog. A frog sitting on a rock "watching" for insects to eat is a good example. The frog does not compute a goal, and it does not correct its current stationary position or the motion of any of its parts. What it does do is respond more or less automatically to the perception of a small moving object by flicking out its tongue to catch the object. Perception is essential to the behavior of the frog.

Thermostat. It seems best to think of the thermostat as a simple man-machine system. Referring to the threefold analysis of computation just given, (i) the goal is set by a person, the desired approximate temperature, (ii) perceiving the actual temperature is done by the thermostat, and (iii) the correction of "trajectory" is made by the thermostat, when it turns the heat source off or on.

Automobile. A better man-machine system of voluntary motion is that of a person driving an automobile. (i) The short-term goal is set by the driver, who may be doing something for another person who sets the long-term goal; (ii) perception also is the driver's responsibility; and (iii) corrections of course are made by man-machine interaction. In a closely related kind of case, that of flying an airplane in bad weather, perception, under conditions of instrument navigation, is also a case of man-machine interaction.

Person walking to store. From the extended discussion of my going to the store in an earlier section, it is clear that this kind of activity is a paradigm of voluntary motion, and all three kinds of computation enter in a direct way. The same may be said of the behavior of an animal chasing a prey, or also of the behavior of an animal trying to escape a predator.

Question of more structure. It needs to be emphasized that the threefold division of computation introduced is not meant to be the end of the story. Each of the divisions introduced has its own complex and subtle structure. The setting of a goal, for example, naturally leads to the introduction of subgoals to provide a structured analysis of what must be accomplished to achieve the desired goal. The perceptual systems of all but the simplest animals are among the most complicated structures known, and much is yet to be discovered and understood about them. It seems to me obvious, and I shall therefore not pursue the point in detail, that the perceptual systems of all the higher animals perform computations of a complex kind, many aspects of which must be learned from experience.

Computation as a natural phenomenon. First of all, the computational power of higher animals evolved over many millions of years from simpler life forms, which in turn evolved even earlier from a soup of amino acids and other mixtures of carbon, oxygen, hydrogen, and the like. This naturalistic view is widely, even if not universally, accepted.

What is not as readily accepted is that computation is as natural a part of our world as the motion of planets or electrons. When I say *computation is a natural part of our world* I mean this in the strong sense that organisms and their computations are just as much an *irreducible* part of the world as is any simple mechanical system. This does not mean that biological organisms and their computational powers are ahistorically given or are in some strong sense universal. In fact their very evolution has surely depended on special environments of unstable energy configurations. The simple planetary systems magnificently studied by Ptolemy, Copernicus, Kepler, Newton, and Laplace could never, as conceived by any of these scientists, been a home for the evolution of life and biological computation. Turbulent environments with the right range of parameters of temperature and pressure, which may be analyzed within either a framework of unstable determinism or of indeterminism, were needed (Suppes, 1991). But any strong thesis of determinism or indeterminism is metaphysically irrelevant to gaining what understanding of these environments and their evolving organisms we are capable of. Almost all of the details are completely inaccessible, and not for historical reasons. The behavior of such

complicated systems could be described only in the most gross fashion even if we were presented with a full panoply of instruments to make observations.

An obvious query at this point is how does what I have to say about the irrelevance of any general deterministic thesis square with the obvious deterministic behavior of a modern digital computer running a particular program. Are, so the query goes, biological computers fundamentally different; in principle, are not they too basically deterministic in character?

Let us deal with the case of the digital computer first. Given a proper stable physical environment with a proper source of energy and a syntactically correct program, the computational steps the computer will take are determined, even if not algorithmically predictable—as seen in the theorem on the halting problem for Turing machines. Of course, this computational determinism does not imply physical determinism of the computer, which is designed with certain physical tolerances in mind. But this issue is not really critical for the main point I want to make, modulo a remark about timesharing or multitasking.

Once timesharing or multitasking computers are considered, even the limited claim of computational determinism is suspect. For a fully loaded system, it is in general impossible to predict the next computational step that will be taken. Even modeling of timesharing as a real-time stochastic process is difficult if not impossible if quite detailed results are wanted.

What is critical is that what the computer will be computing tomorrow or the day after or next year is completely dependent on what program input it receives, and as our analysis of voluntary motion shows it is hopeless in principle to predict on physical grounds what program inputs will be received in the future. Moreover, there is no strong ground for believing a wider sense of predictability based on psychological and social concepts as well as physical ones has any possibility of being successful. Detailed predictability or determinism has no working place in such an extended study of computer use. Either concept is as out of place here as either would be in a year's data on the movements of a monkey, dolphin, or some other animal in the wild.

Given an appropriate physical environment, the computer remains ready to compute when it is given program input. It is a transcendental metaphysical fancy to think that its computational behavior over a period of a week or of a year can be analyzed without remainder in deterministic physical terms. The central reason is that the system is an *open* one. It is not possible to isolate the system because program inputs are continually arriving from a variety of sources for a variety of purposes. These inputs are like stochastic stimuli impinging on the peripheral nervous system of an animal and why, for the same central reason, animal behavior is that of an open system. Even purely physical systems that are open cannot be analyzed in detail either from the standpoint of determinism or of predictability.

What I have said about digital computers being open systems and therefore unpredictable applies in even stronger terms to biological organisms, which have evolved to survive in a much greater variety of environments and stimulus inputs. Biological organisms are necessarily open systems. As open

systems they must make computations about their environments in order to survive. These computations, concerned in the case of voluntary motion with goal-setting, perception, and course correction, cannot be analyzed away either in terms of a sequence of deterministic causes or in terms of a sequence of probabilistic physical causes. Not only understanding but predictability requires use of the concepts intrinsic to voluntary motion.

Computation by biological organisms evolved along with other central features of life as essential to survival. The gist of the argument I am making is that free will as exhibited in voluntary motion is biological rather than merely moral in nature, and is essential to the survival of organisms.

The rational and moral tradition. In contrast to the biological view of computation and free will I have been arguing for, the greater part of the philosophical tradition has put the weight of the argument for free will elsewhere. One line of argument has been that the introspective evidence of deliberation is proof of the existence of free will in humans but not necessarily in other species. I see no reason for such a special focus. Deliberation is a form of computation, and what is biologically fundamental is the widespread role of computation in the continuing struggle to survive. Sometimes the argument about deliberation is given as one about rational belief, but this can only be taken as a further constraint on computation, not as a substitute for it.

The dominant philosophical tradition has been to make the existence of free will a necessary condition for the existence of moral responsibility. It is this tradition I want to examine again. The first point to note is that making free will a necessary condition of moral responsibility is a problem for the viewpoint I am advocating *only if* this relationship is made the primary focus of arguments for free will. But this is what is often done. The fact of the existence of moral responsibility in humans is used as the main—or sometimes, the only—argument for free will.

Although he certainly had predecessors, it was especially Kant who occupied the high ground of morality in his defense of free will, and I want to concentrate on what seems to be the central even if implicit assumption of Kantians and non-Kantians alike. This is that moral responsibility entails a kind of freedom that goes far beyond any biological account of the voluntary behavior of animals. The kind of freedom implied is one that derives from a theological rather than a scientific tradition. Again, it is to be seen in Kant, who presents as the great triad of problems for philosophy: the existence of God, the immortality of the soul, and the freedom of the will. Another but quite related way of expressing this view is that among the animals only humans have a moral conscience.

Admittedly the explicit theological source of this "high" ideal of freedom is often suppressed. In the tradition of much modern analytic philosophy, the argument that moral responsibility entails free will is made on as narrow a ground of common sense and moral psychology as possible. It is characteristic of these arguments to be very nonbiological. The behavior of other species is not seen as relevant to the argument. Indeed, it would not ordinarily be thought

of as part of philosophical psychology to recognize the biological origins of the human species.

Putting the matter this way raises an important methodological issue that bears directly on the problem of the source of freedom. I have taken a biological line of argument that makes free will a natural concept exhibited in the behavior of other species. This is clearly a scientific line of attack. Those who center the argument for free will around the existence of moral responsibility implicitly or explicitly assume (or believe) that the grounds of the argument are philosophical rather than scientific.

It is also a feature of most of the purely philosophical arguments for free will in contemporary philosophy that the arguments are quite simple in structure. To a large extent they rest upon the premise that moral responsibility exists, and if an argument is wanted to back up this assertion it is that moral responsibility is a necessary concomitant of being a human being.

These facts or assertions I grant, and they are not the focus of dispute. What is in dispute is whether this kind of philosophical argument resolves the puzzle about free will and determinism. The proper philosophical focus, so I would claim, is to analyze away the apparent conflict between determinism and free will.

The resolution I have given should be evident, but I recapitulate the main strands of my argument, some of it given in Suppes (1993).

I. Determinism does not imply predictability.
II. Determinism and randomness are consistent.
III. Universal determinism or universal indeterminism is a transcendental metaphysical assumption.
IV. Voluntary motion has three characteristics:
    1. Physically unpredictable;
    2. Predictable given knowledge of goal;
    3. Perception and path correction required.
V. Voluntary motion is characteristic of many animal species.
VI. Voluntary motion requires several kinds of biological computation.
VII. Biological computation has evolved as part of the struggle to survive and is a natural part of the universe.
VIII. Biological computation exhibits free will.
IX. The puzzle is resolved because there is no scientific theory of universal determinism (or indeterminism).
X. There is no transcendental nonbiological origin of moral responsibility.
XI. Moral responsibility is part, but only part, of the natural history of free will.

NOTES

1. This idea was already clearly expressed by Aristotle in *Movement of Animals*, 701b25.

2. One familiar line to take that is congenial to the framework I have been setting up, but does not go far enough for my taste, is this: actions are free, and therefore exhibit free

will, when they are not completely constrained by external causes. Here the problem is to clarify the idea of complete external causation. Constraints, of course, are the order of the day, even for actions that we think of as typical examples of being free. The movement of my hand, for example, is constrained by a great deal of obvious physics, such as the maximal rate of acceleration, the maximal rate of deceleration, the rate of final velocity, etc. But these constraints from physics are only partial constraints. There is much freedom left after they are satisfied—a point to be amplified later. On the other hand, if I grab your arm and force it to move without your consent, this is scarcely an example of free action. This ground has been tread before and I do not really want to retrace a path across it one more time.

REFERENCES

D. R. Holland. 1993. *The Quantum Theory of Motion: An Account of the de Broglie–Bohm Causal Interpretation of Quantum Mechanics*. Cambridge.

P. Suppes. 1985. "Explaining the Unpredictable." *Erkenntnis* 22: 187–95.

P. Suppes. 1991. "Indeterminism or Instability, Does It Matter?" In *Causality, Method, and Modality*, edited by G. G. Brittan, Jr. Dordrecht.

P. Suppes. 1993. "The Transcendental Character of Determinism." *Midwest Studies in Philosophy* 18: 242–57. Notre Dame, Ind.