Volition: a Semi-Scientific Essay

By William T. Powers

Every child wonders, sooner or later, how it is that simply wanting one's hands, arms, legs, body, head, or eyes to move suffices to create the wanted result. The sense of willing that one's own body do something is at the same time unmistakable and unexplainable, being unlike any other mental or physical experience.

While willing an act seems to suggest that we are masters of our own behavior, experiences of other kinds suggest just the opposite. As children we do as we will when it is playtime, but from the very beginning we find that we also do as we must, when people and events decree that playtime is at an end. Even the passive physical world forces us into action in ways that can seem to push the will aside. With growing force, necessity makes itself known in many forms. The demands of our bodies, saying that we must breathe, eat, drink, stay warm, seek love, and avoid pain, override the will more and more often; one demand leads to another, until by the time we are adults it can seem that we no longer have any freedom to will except as a momentary act of useless defiance. When the rat-race is at its worst, there seems to be an external reason for every slightest act from rising in the morning at the alarm clock's buzz to swallowing the final nightcap so we can sleep, only to rise, too soon, again. To indulge in any extended period of purely volitional action would be to put unacceptable stresses on the network of behaviors we are forced to adopt, stresses that seize control again and bring us back into the daily groove, will we or nil we.

The transition from childhood to adulthood is unpleasant largely because of the sense of steadily diminishing freedom to will. On the one hand, adulthood promises immense freedoms—driving a car, getting out of school, having one's own money, going to bed when one pleases, being listened to, understanding how things work, owning and managing things and events. On the other hand, adults obviously do not seem to enjoy these freedoms as much as they ought to. In fact, they seem to act as if they have no great amount of freedom. Every child must at some time vow not to become like that—not to give up control of one's own life. And every child inevitably ends up breaking the vow, perhaps raging but in almost every instance succumbing to all the controlling influences that prove unavoidable.

The traditional scientific view of behavior is the adult's view, not the child's. But this is not the view of a wise adult; only of an adult who has decided that the sense of will that was given up must somehow have been an illusion. Opponents of the objective, dispassionate analysis of causation that is traditional in science, on the other hand, maintain the child's view, insisting on the essential freedom of the mind with a child's faith—with the same amount of influence on science that children usually have on adults.

The puzzle of the will is central in our attempts to understand human behavior. Insisting that creative will is all, the naive child's view, is neither more nor less correct than insisting that it is impotent or nonexistent, the cynical adult's view. To understand both will and necessity, we must avoid siding with either view, and try to define the terms of this puzzle in a way that gives us a chance at solving it.

Internal vs. External Causation

To speak of volition as a sense of willing is to use one word in place of another, illuminating nothing. While only the individual can sense volition when it is occurring, the ability to sense it confers no particular understanding of it. If sensing it were enough, we would not have these problems. What we must do is find a place for volition in our general understanding of both private and public, but most importantly public, phenomena.

Volition can be defined as a cause of behavior that is internal to the behaving system. Speaking

Note: This article originally appeared as Chapter 2 in W.A. Hershberger (Ed.) (1989). *Volitional action: Conation and control*, Amsterdam: North-Holland. Reprinted by permission of the author.

generally instead of personally, we can see that human behavior seems to have two kinds of causes. One kind we can easily see, as when a gust of wind makes a man struggle to stand up, or an unexpected sound makes someone jump, or a worker tries harder when the boss threatens to fire her. The other kind is harder to see, because the cause is located where it can't be observed; the identification of a volitional act always seems weak because all we can say is that there was no apparent external cause. Few of us would dare to claim that we have noticed every possible cause and ruled it out. The weakness of the identification would seem to leave external cause as the most rational choice.

On the other hand, a careful consideration of human behavior, our own or that of others, makes it quite clear that we cannot identify many external causes. While we can pick out salient events such as an explosion or the offer of food and make a case that the ensuing behavior was the result, it is much harder to extend these connections to all behaviors and all events. Given any event chosen from the ongoing stream at random, we normally have no way of predicting what behavior will follow it in any given person. And if we really pay attention to behavior, we must admit that behavior is going on every moment of a person's life, in an unending continuous flow. It isn't just that our knowledge of external causes of behaviors is incomplete: it is nearly nonexistent. In sheer quantity, the amount of behavior that has been connected to prior causes is only an infinitesimal fraction of all the behavior that goes on every day.

Scientists who have given up completely on internal causation have done so not because of the evidence, but because of an urge to simplify. It is much easier to assert that all behavior is externally caused than it is to envision trying to sort out one class of causes from another. In support of external causation, it has been claimed that in a physical universe, all material objects are caused to behave by the confluence of all current influences on them. But this physical principle is not a premise from which we can conclude that all behavior is externally caused: it is simply a restatement of the assertion in different words. And it is a restatement that ignores all the ways in which organisms differ from the simple point-masses to which the original Newtonian principle was applied.

The principle difference is complexity: there is a great deal more going on inside an organism than inside any piece of matter that a physicist or a chemist studies. Most pieces of matter that a physicist studies do not stand up and try to get away. This complexity means not only that there are important processes going on inside the organism at all times, but that these processes may arise from sources that existed at unknown and unknowable times in the past. When a person speaks, the grammar and syntax that shape the speech may have originated in the outside world, but they certainly did not originate just before the utterance. As far as any present-time observer is concerned, the causes of grammar and syntax *now* are carried in the brain of the speaker, and cannot be traced to anything happening in the current environment.

This gives us the first wedge with which to pry open the puzzle. We must admit at least that large parts of the behavior we observe have origins that are unconnected with the current environment. On this basis we may still claim that some behaviors amount to responses to current stimuli, but we must allow that even more of the behaviors, perhaps most of them, must be under control of processes that are not altered by any present stimuli.

The Logic of Causation

Another route we can take involves a closer look at what is supposedly caused, behavior itself. What is behavior? The naive view, which is shared by scientist and laymen alike, is that behavior is whatever organisms do. But what do we mean by "doing?"

The two-letter word "do" takes up four column-inches in an old Collegiate dictionary, and far more in an Unabridged. This primitive grunt refers to causing essentially any occurrence that can be named. It asserts agency, but reflecting our ignorance it skips over process. The doer does, but how the doer does it is not mentioned.

How does one open a door? Not, we can be sure, by opening it -that is not an answer to a "how" question, but a reassertion of agency in more obscure form. Normally, one opens doors by pushing on them or pulling on them. Opening a door would surely be classed as a behavior, but in fact this behavior is carried out by an organism that is doing something distinctly different from "opening." It is the door, not the organism, that opens. What the organism does is to apply a force with its muscles: the consequence of this effort is, usually, that the door opens.

Most of us will open a sizable number of doors in one day, some familiar and some unfamiliar.

We open bedroom doors, bathroom doors, front doors, car doors, supermarket doors, refrigerator doors, cupboard doors, and the doors where we work. There is no linguistic problem with calling all these activities "opening doors," but in terms of the motor actions we carry out, not only are the actions very different over all these instances of the "same behavior," but they are quantitatively different each time we open the *same* door.

What we call behavior is really some repeatable recognizable consequence of our motor actions. Almost 100 years ago, William James pointed out the uncomfortable fact that while these consequences repeat, the actions that bring them about do not repeat. Had James gone on to analyze this observation in more detail, he would have realized that the actions do not repeat for the simple reason that if they *did* repeat, their consequences would vary. If you turn left to enter a cafeteria, you will be able to buy your lunch. But if you enter the building by a different door, or if someone is standing in the way, or if the cafeteria is locked, you will not get lunch by turning left. Something else will happen. This is the story of essentially every behavior of any amount of complexity. Circumstances change. The surrounding world influences the outcomes of actions, and those independent influences can change greatly from moment to moment. Sometimes they don't change, so the same action will have nearly the same result as before. But organisms must produce behavior in the worst-case world, too, and they do. When external influences change, organisms alter their actions to compensate, even to the extent of reversing them or substituting a totally different action. This is a commonplace fact of life: regular behavior is *not* brought about by regular motor actions, and regular motor actions would not normally produce regular results.

That fact, as simple and obvious as it is, spells great difficulty for the concept of external causation. For external causation to work, the causal chain must remain predictable from beginning to end. There must not be any other causes that contribute to the outcome downstream from the initial cause—otherwise, anything could happen. If the principle of external causation worked as it is supposed to work, we would predict that disturbing the outcome directly would cause the outcome to change, in exact proportion to the disturbance. What does happen is that an immediate change in the action just cancels the effect of the disturbance.

This is the only way in which organisms can possibly continue to produce recognizable behavior. The patterns that result from their motor actions are ordinarily under continuous disturbance, the disturbances arising partly from independent sources in the environment and partly from the varying relationships of the organism to its environment. We see stable patterns; it follows that the actions of the organism cannot be correspondingly stable.

This analysis would seem to rule out external causation altogether, but that is not quite the result. What happens instead is that we are made to focus on something outside the purview of the causal hypothesis—not what changes when stimuli and disturbances occur, but what does not change.

The Logic of Control

We do not normally pay attention to the motor acts by which familiar patterns of behavior are created; for one reason, they are hard to observe. It is not action, but the consequence of action, that is made to repeat by a behaving organism. To understand just how variable those acts must be, we have to understand something about the workings of the physical world. When we see a person reaching out toward the floor-selector button in an elevator carriage, we see what seems to be a motion of the arm directed by its muscles toward the button. With a little reflection, we realize that the muscle forces are not aimed in the direction the hand is moving. They are aimed primarily straight up, countering the force of gravity and whatever accelerations of the elevator carriage are occurring. Even an act like reaching out toward something, which seems a direct expression of muscle action, is several steps removed from the actual motor behavior that is going on. The ends and the means are almost never related in any simple straightforward way.

Clearly, we do not simply "do" behaviors. That description is just too sketchy. A more accurate description would be that we—and other organisms—act in such a way that certain consequences are brought about and maintained. The phrase "in such a way" has a specific meaning: the way in question can be deduced from observing the consequence and knowing what independent forces are acting to alter the consequence. For instance, if we observe a car being steered straight down a level flat road, and we know that a crosswind is exerting

75 pounds of force on the car to the left, we can be quite sure that the driver is exerting a force on the steering wheel that, relayed through the power steering, the linkage, and the front tires, pushes the car to the *right* with a force of just 75 pounds. If that were not so, the car could not go straight. When more than one influence adds a sideward force to the car—the camber of the roadbed, for instance, adding its effects to those of the crosswind—we can be quite sure that the driver's effort, translated into an effect on the car, is equal and opposite to the sum of all those disturbing forces. That is simply a matter of applying Newton's laws of motion, and observing that the car continues in a straight line.

When we see consistent behavior in the presence of independent disturbances, we can deduce that the actions of the organism must be varying so that the resultant is right for producing what we see. This is the basic logic of the phenomenon we know as control. A disturbance that tends to alter the final pattern results immediately in a change of motor action that tends to alter it by the same amount in the opposite direction. The net result is no change, or almost none. It is this lack of change, under circumstances where change is to be expected, that tells us control is occurring.

This concept of behavior clearly does not fit the conventional causal model. As expressed so far, it seems to rely on variations in actions that are fortuitously just right to prevent disturbances from having disturbing effects. To implicate external causation in this kind of situation, we would have to imagine that the external cause varied in just the way needed (taking the organism's properties into account) to make behavior change to preserve a particular outcome. We would have to imagine stimuli that act on the driver so as to keep the car exactly in its lane for, say, 100 miles despite the myriad disturbances, mostly invisible, that come and go during the trip. But the driver's environment doesn't care whether the car stays on the road or goes wandering off among the sagebrush. The causal explanation requires us to believe not just in one incredible coincidence, but in a never-ending stream of incredible coincidences.

To define behavior as a process of control does not require us to explain how this process is brought about: first we define the phenomenon; then we try to understand how it is created. The phenomenon is this: by varying their actions, organisms stabilize certain outcomes of those actions, outcomes that would otherwise change with every change in environmental influences on the same outcome.

The Mechanism of Control

The development now turns somewhat technical. The question before us is now how an organism must be organized to produce the control phenomena we observe. The answer to this question has, in fact, been known for some 50 years.

If independent external causes cannot account for behavioral changes that control consequences, we must look for the causes elsewhere. The solution of this problem was found by engineers who studied certain types of human behavior in order to replicate it in a machine. The resulting machines were called control systems. The missing factor, these engineers discovered, was that the control system must sense the very consequence or outcome that is to be placed under control. The external cause of control behavior is the outcome itself—the effect. The cause and the effect are identical. The cause is not independent of the effect.

The basic arrangement of a control system is simple. A sensor reports the state of the controlled variable as a correspondingly variable signal, inside the control system. This signal is compared against a reference signal carried inside the system, and the discrepancy is represented by still another signal, the error signal. The error signal is amplified to produce a physical output, which in turn acts on the same controlled variable. This is the famous feedback loop, the feedback being negative in that any change anywhere in the loop propagates all the way around the loop to arrive at the starting point with the opposite effect. Properly speaking, feedback is a property of the entire closed loop, not of any one part of it.

When such a system is properly designed (not a particularly difficult task if the system is simple), the result is not quite what may have been expected. The basic effect is that the sensor signal is held very actively in a match with the internal reference signal. If the controlled variable is disturbed, the beginning of the change due to the disturbance causes a slight departure of the sensor signal from the reference signal; an error develops, which, highly amplified, produces action. The action, simply because of the way the negative feedback loop is arranged, tends to force the controlled variable back toward its undisturbed state, and thus tends very strongly to force the sensor signal back toward a match with the reference signal. Almost as an afterthought, this action opposes the effect of the disturbance.

The only generally correct way to describe the action of a control system is as a system in which all influences are in continuous equilibrium all around the closed loop. Applying a disturbance to the controlled variable results in an immediate rebalancing of the equilibrium, the action changing as the disturbance changes, so that the sensor signal is never allowed to depart much from the setting of the reference signal. Intuitively, we want to think of this circle as a sequence of events going around and around. Intuition, in this case, is simply wrong: it is attempting to treat the closed loop as if it were a lineal temporal sequence, and that does not work. Only the mathematics of control theory (or hands-on experience with control systems) can show the essentially simultaneous action of all parts of the system. Intuition must be retrained.

If anyone's intuition objects to the idea that mathematics can help it, the proof that it can is to be found in a basic property of control systems called "loop gain." Loop gain is the amount by which any variation is amplified as its effects make one complete trip around the " loop. Real control systems normally have loop gains amounting to a factor anywhere between 10 and one million. In other words, the effect of a small change in a variable upon itself (via the closed loop) is a change from ten to a million times as large as the original, and in the opposite direction. Intuition, of course, predicts disaster. Instead, there is control. One must simply learn control theory to understand how this result can occur. Nothing in our intellectual training has prepared any of us to reason out, unaided, how control systems work. The principles involved, although 50 years old, are unknown to almost everyone but engineering specialists.

Using the principles of control theory, engineers have built machines that behave exactly in the way organisms behave. They automatically vary their actions to bring about and maintain specific predetermined consequences of those actions, counteracting disturbances without any specific instructions to do so. They produce consistent outcomes by variable means: they behave just as William James said organisms behave. That is no coincidence: they were modeled on the behavior of organisms, and the engineers who invented them succeeded, serendipitously, in finding the first workable model of a behaving organism.

The Appearance of Control Behavior

When an engineer builds a control system, providing a reference signal for it is just a matter of introducing a signal generator into the system. The source of reference signals in organisms is not quite that easy to explain, but we do not need to account for the presence of reference signals to understand their effects.

For all practical purposes, reference signals function exactly as intentions are supposed to function. The reference signal specifies an intended state of the sensory input. Action is based at all times on the difference between the sensory input and the reference signal. The action, having a polarity opposite the detected difference, serves to reduce or negate that difference. This negative feedback first brings the external variable to the specified state, and then keeps it there, all the while creating actions that oppose any disturbances that might also act on the variable. Thus completely without any predictions and certainly without any influence of the future on the present, the control system's reference signal determines the outcome of action.

The action of a control system makes its sensory representation of an external variable match its internal reference signal. If that internal reference signal changes, the same organization will force the sensed variable to change in the same way, maintaining the match between sensory representation and reference signal. Thus whatever can vary the reference signal can cause the external variable to vary in the same way. The behavior of the external variable is then no longer what it would have been with the control system, the organism, absent. Normal physical influences are treated as disturbances, and cancelled by variations in the output actions of the control system. The external variable affected by the action behaves as the reference signal specifies, not as the environment otherwise would make it behave.

Reference signals clearly have something to do with the phenomenon we intuitively recognize as volition. The simple alteration of a signal inside the system causes an external variable to behave in a corresponding way. But this causal connection is anything but straightforward, because the motor outputs that appear not only must bring the variable to the right state, but must show added variations that are needed to counteract the effects of unpredictable disturbances. In a great many situations, the outputs required to keep a variable under control are small and even trivial—or would be, if disturbances were not present. Disturbances, however, are almost always present, and even in perfectly normal environments they have large influences on the variables we are controlling. A driver in a precisely-made car in perfect condition on an absolutely level road would scarcely need to steer at all—the efforts involved would be miniscule. But if the road tilts and the crosswind blows, or if the car pulls spontaneously to one side, the driver must start exerting significant efforts, efforts that are needed simply to oppose disturbances. Because these efforts do occur, the controlled variable is kept from changing; it obeys the intention, not the disturbances.

The logic of control shows us that there are really two major kinds of relationships going on at the same time. One is the relationship between the reference signal and whatever it is that is being controlled. The behavior of the reference signal determines, through feedback effects, the behavior of the controlled variable. At the same time, however, there is another relationship between the system's actions and independent environmental disturbances. Every disturbance calls forth a change of action that is quantitatively equal and opposite to it, in terms of effect on the controlled variable. We know that this apparent relationship is really the result of small errors induced by the disturbances, errors that are highly amplified to become opposing actions. If we did not have that model of a control system in mind, the appearance would be that the disturbances are directly causing the actions, and the stability of the controlled variable would be just a lucky break for the organism.

We therefore have a dual causal relationship that is seen in the behavior of every control system. The *actions* of the system appear to be determined largely by external forces that disturb the controlled variable. At the same time, the state of the controlled variable appears to depend only on the will of the control system, which we now recognize to mean on the setting of a reference signal inside the system. The controlled variable remains close to the state specified by the reference signal. We see the arms of the driver urging the steering wheel continuously to the left and right in an apparently random pattern, a pattern we could eventually trace to crosswinds and other variable influences on the car. But the car itself continues its course undisturbed, remaining on the line that the driver intends. What the car is doing seems to be almost unrelated to what either the crosswind or the driver's arms are doing.

These two seemingly different kinds of causal relationship are really just aspects of the way one system behaves in relationship to its environment. Control theory removes the duality, showing us what is really going on. But while it does that, it also explains why we see two different kinds of causation in behavior, external causes and, less obviously, internal causes. The reference signal is the internal cause, and what it causes is the outcome of behavior. The sum of all disturbances is the external cause, and what it causes is the action, or most of the action, that stabilizes the outcome.

Control theory thus shows us how it is that outcomes can be voluntary while actions are involuntary (a nice summing-up that is due to Wayne Hershberger). Once we have this picture clear, we can understand how the driver can intend for the car to stay on the road, and carry out that intention, while being unable to predict or choose the forces his own muscles apply to the steering wheel while bringing about the intended result. When the driver elects to control the position of the car, by that very choice he elects to let the wind and a dozen other invisible disturbances determine his motor actions.

A Hierarchy of Control

Motor behavior involves the operation of hundreds of control systems, each associated with controlling the force applied at the attachments of a muscle. Many others sense and control muscle length. But these elementary control systems are not the end of the story: they are used in turn by systems of higher level, which control variables much farther removed from the nervous system. In the example of the driver, the muscle-force control systems are employed in the larger control loop that involves the steering forces applied to the car, the position of the car on the road, and the visual images that tell the driver about that position. In order to control the appearance of the scene in the windshield, the driver's primary way of sensing the car's position, the driver's brain must compare the scene as it actually is with a reference image (or, if not literally an image, some internal information relating to the visual field). The mismatch between what is sensed and what the internal reference specifies is the basis for exerting forces to the left or right, or for not exerting forces on the steering wheel.

The higher-level control loop does not operate the muscles directly; instead it varies reference signals sent to the muscle- force controlling systems (according both to this control-system model and to neuroanatomy). Those control systems automatically make the sensed forces match the reference signals, in the process generating physical forces on the steering wheel. There are probably more than just these two layers of control involved in steering a car, but these two will get us started.

The reference signal specifying the car's intended position is itself variable: the driver is not stuck forever in his lane. When the driver overtakes a slower vehicle, we observe that at some point the car veers left and takes up a new path in the adjacent lane until the vehicle is passed; then it swings back and resumes its former position. In a stiff crosswind this can be an exciting encounter as the car passes into the lee of the other vehicle; at that point the steering effort that has been counteracting the crosswind suddenly makes the car lurch toward the other vehicle, and the steering effort has to be relaxed—and then proves insufficient as the driver's car pulls ahead, into the crosswind again. But most drivers manage to pass another car or a truck in a way that seems effortless to an onlooker who does not feel the fluctuations in steering efforts.

This passing-event required that the reference position for the visual-motor steering control system be changed for a while, and then changed back. But following the logic of control, we do not ask so much about these changes as about what remained constant because of them. What remained constant was the car's progression toward its destination. There is no one generic answer to the question of what remains constant—the driver might be trying to maintain a constant estimated time of arrival, or might just be trying to maintain a good average speed for some unexamined reason. Keeping the speedometer at a certain reading would be part of maintaining an average speed, but going around a truck instead of ploughing into its rear is also necessary.

Voluntary and involuntary aspects of the behavior shift their roles as we consider higher levels of control. If the driver chooses to exert a specific sensed force on the steering wheel, he has no choice but to create a certain amount of contraction in his muscles. If he chooses to keep the car in a specific position on the road, he has no choice but to set the muscle-force reference signal at whatever level is required by disturbances of the car's path. In effect the crosswind and other disturbances are determining the setting of the effort reference signal, given the intention to stay in the lane. And now the intention regarding the car's position relative to the road has to be changed if the forward progress is to remain the same: the presence of the other vehicle makes the changed position mandatory, given the intention to maintain forward progress.

Again we ask, what is this forward progress for? Presumably, the driver is not astonished to find himself driving a car down a road: he is going somewhere, perhaps intending to arrive in time to meet someone for lunch. The intention of arriving at a particular place at a particular time has put him on this road, in this car, going at this speed. However, if the perception of arriving in space and time as intended is to be maintained, the reference signal specifying forward progress has to be varied: it must have varied in order to get the car onto this road in the first place, and sooner or later it must vary in order to enter the driveway of the restaurant. To maintain the pattern of the whole trip in the intended form, the driver must periodically vary the intention regarding forward progress, and in the precise way dictated by the starting point, the time on the dashboard clock, and the location of a free parking slot at the destination. The reason for having made and now having kept this lunch date is for the driver to sell a house to the person waiting for him. The driver intends to sell this house. If someone else had called him to ask about it, he would have made a different trip, perhaps not even in a car, and he would have gone to a different destination, perhaps not for lunch. That is because once he has selected the reference condition of selling a house, he has to go wherever a buyer can or will meet him. There is no other way to give his pitch to the prospective buyer: he has no choice.

As it happens, our driver was trained as a physicist specializing in nuclear power plant design. Why is he so intent on selling this house? And why was he so intent last week, and why will he be the same next week? Because selling houses is now his only means of making money, the demand for new nuclear power plants having slackened dramatically. This means of making money presented itself, and as he intended to make a reasonable living and no comparable opportunity was found, he had no choice but to take the job. This was the only available occupation that promised to provide the amount of money he intended to make.

The intention to make $50,000 per year instead of, say, $25,000, can be traced to the fact that when he lost his job as a power plant designer, he consoled himself in a foolish manner and is now required to pay $25,000 per year in alimony. Actually, his simple needs would be met quite well on $25,000, but the negative $25,000 disturbance due to the alimony required him to set his salary goals correspondingly higher, so he can net enough to provide a sufficient living for himself. Obviously, he intends to make a sufficient living, as he thinks of it, but that intention, plus the disturbance, leaves him no choice but to earn twice as much as he needs.

We can now see that it is the alimony disturbance of the driver's income that explains why, at 11:48:37 this morning, he was exerting a 1.2 kilogram-meter torque to the left on the steering wheel, steering the car to the *right* around a curve in that ubiquitous crosswind. The highest-level goal—plus dozens of external disturbances at several intervening levels of abstract intentions—required that effort at that time.

If we were to carry out this sort of analysis with a real person, we would arrive eventually at levels of intention that would be very hard to trace any higher. Perhaps there is a highest level, having to do with control of abstract concepts like a self, relationships to a society or a family, loyalties to knowledge or culture or religion. Where the highest-level reference signals come from is an interesting question, but not germane here.

The central point of this imaginary excursion up the levels of control is that volition and necessity are not simple matters. It is rather arbitrary to select a momentary intention and treat it as if it came from nowhere and served no higher purpose. It is especially risky, in talking of the will, to talk of *free* will. What seems free will at one level of analysis is a necessary adjustment to external disturbances at another level. There is nothing wrong with identifying the sense of volition with reference signals in a hierarchical control-system model of the brain. That may well be a correct identification; it is certainly functionally and scientifically plausible. But in order to understand how voluntary and involuntary behavior interact, we must think of the entire hierarchy, not just one slice out of its middle.

The Web of Intention

Even at the lowest level in the human behavioral hierarchy there are control systems, systems that maintain muscle forces, as sensed in the tendons, at levels specified by signals descending the spinal cord from the brain. Those descending signals, while acting as first-order reference signals, are also the actions of higher-level control systems concerned with controlling more abstract or general variables.

There must be many major levels of control, perhaps ten or more, in the human nervous system and brain. At the lower levels we have systems that sense and control effort vectors in space, that employ these vectors to control bodily configuration, that vary configuration reference signals to control movements or transitions. At still higher levels the configurations and movements become the familiar events we recognize as acts, and those acts are maintained in relationships involving many acts and many external objects and events. On top of these levels are all the cognitive levels, in which the world of experience is classified, analyzed symbolically and logically, abstracted to become principles and generalizations, and finally made into coherent concepts like the concept of a self, a society, a science, a material world. Control occurs at all of these levels, each level acting to control its own kind of perception by means of varying the reference signals, which we experience as volition, reaching lower systems.

While it may be that human beings control what they experience in terms of certain broadly shared types of perception, the variety of human experiences, circumstances, preoccupations, and problems tells us that within these broad classes, the structures of control that individuals build up as they mature are highly idiosyncratic. It is no simple matter to manage a world that begins as millions of identical sensory signals, and is then subject to multiple levels of interpretation that must, for the most part, be worked out in private and without the aid of an instruction manual. It is no simple matter to discover how one part of this world can be controlled without negating the control of another part of it, at the same or a different level. The high-school senior understands that by going to college and submitting to at least four more years of school, he will be able to enhance his personal power and self-respect, to raise children in comfort, to feel a part of his conception of a larger world.

But if he chooses that intention, he will have to tolerate continued supervision by his parents and others, he will have to leave behind the girl he loves, and who will take care of his cat?

The loss of volition sensed by the adolescent—and many who are much older—is not really a loss of volition, but a gradually expanding network of self-contradictions, a consequence of ignorance about how we work. The physical world and the society into which we are born only set the stage on which our lives are played out: they do not limit our freedom, but simply constitute the means available to us for doing whatever we can make sense of doing. It is up to each of us to learn how to act on and in that world, to learn to perceive its possibilities, and to learn how to organize our intentions regarding that world. Through the miracle of communication each person can learn from the others, but if there are no others who understand human organization, the amount of help available is going to be small. People are very free with advice, but as advisors tend to contradict each other, the useful residue is not as useful as it might be. Look before you leap—or nothing ventured, nothing gained?

Beneath the fuzziness of personal experience there lurk some hard natural laws. The process of control itself, at any level, requires that certain mathematical relationships in space and time be properly established. Fortunately we seem to have the capacity to reorganize until we achieve skillful control. But there are even harder laws. Given a body containing about 800 muscles (depending on how they are counted), it is mathematically impossible to establish control of more than 800 independent variables of experience at the same time. The degrees of freedom of control cannot exceed the degrees of freedom to act. And actually to be able to control that many variables at once, one would have to solve 800 nonlinear differential equations in 800 unknowns. It is unlikely that the nervous system—even the nervous system of an engineering mathematician—would be able to realize anything near that potential.

And that takes into account only the second level of control. Now we must consider that the variables of the second level, already abstracted once from raw sensory inputs, are abstracted again to yield a new type of experience, and thus a whole new set of potentially controllable experiences. And this adding of new modes of control at new and ever more abstract levels must continue for at least some respectable number of levels. In every case, at every level, the same mathematical problem exists: how to partition the universe of experience so that its parts can be independently controlled without self- contradiction; without conflict.

This whole hierarchy of control contains a network of intentions that represent the actions of the control systems above the first level. When, inadvertently, the intentions cancel each other before they can produce any action, we feel a loss of volition, a paralysis of the will. At the highest levels our intentions are reasonably clear, but at the lower levels they may demand contradictory intentions, and so produce none at all, or only an unsatisfactory compromise. We easily become lost in the complexities of managing this physically compact but functionally gigantic structure, the human brain.

How many of us could sit down and draw a map of our structures of intentions? Most of us could probably explain fragments of the structure here and there: this act serves that purpose, which in turn was selected as part of satisfying a higher-level intention, and so on for perhaps three or four levels at the most. A few of us might be able to show how the goals we seek at work relate to those we seek on weekends, or how our relationship to our parents interacts with our relationships to our wives and children. It is unlikely that any person alive could draw the whole map, even considering just the parts of it that are actually available to inspection. When we consider our own lives, we see them as if through a moving peephole that limits the size of the picture visible at a given moment, or as if we are shining a penlight around in a dark cathedral, trying to build up a picture of the whole huge room out of images that pass through the small circle of light.

The sciences of life, being founded primarily on the old causal model, have little to tell us about understanding the vast structure of the mind. Having long ago dismissed the importance of phenomena such as volition, they have produced essentially nothing that would help us to map out our own organizations, either to understand or to improve them. Control theory, on the other hand, seems to show us the way toward doing something useful in this direction.