Chapter 16

Free Will

16.1 The Problem

The nature and existence of free will are traditional philosophical topics. One can easily see why by noting that the notions of free will and moral responsibility are related. The nature of the connection is a matter of debate; however, one can minimally say that paradigmatic cases of moral responsibility occur only when the agent acted of his own free will. Many have viewed free will as threatened both by theology and science.[1] Leaving theology aside, the main threat coming from science has been represented by determinism, which we must now briefly consider.

Determinism is the view that if at a time two isolated systems are described by the same state equation, then they have always been and will always be described the same equations and that two systems in the very same state always display the same properties. In effect, determinism is equivalent to the combination of evolutionary and value determinism. Classical mechanics is often taken to embody determinism, but to what

---

[1] The theological problem is this. The traditional orthodox God knows the future. If God knows today that tomorrow at 3:23 pm I shall mow the lawn, then it looks as if mowing the lawn is not really up to me, since if I were to abstain from it I would turn God's knowledge into mere belief, something that both philosophy and piety seem to prevent. Note, by the way, that the problem, albeit in a slightly different version, remains even if we dispose of God. Today it is either true or false that tomorrow at 3:23 pm I shall mow the lawn. In either case, mowing or not mowing is not really up to me, or so it seems. These are interesting issues, which however we cannot address here.

extent such a view is correct in all cases is far from clear. In well-behaved physical systems, solutions to Newton's Equation of Motion exist and are unique, and problematic cases are usually solved by appealing to physical intuition (Symon, K. R., (1960): 23-4). Nevertheless, it is possible to concoct situations in which either the existence or the uniqueness of solutions does not obtain.[2] However, as often such situations appear to involve excessive simplification of the actual physical conditions, and in addition they are typically very artificial, we may consider them marginal to our topic.

One of the commonly held intuitions allegedly associated with the notion of free will is that of the capacity to do otherwise than one did. If one did something of one's own free will, then one could have done otherwise: one had a real choice, as opposed to merely believing one had it, and one chose what one did among a number of available options. I ate vanilla ice cream but I could have eaten chocolate or nothing at all. Some philosophers, the *compatibilists*, gloss on such a capacity as follows. Saying that I could have done otherwise really amounts to saying that I would have done otherwise had I decided to. When I say that I could have eaten chocolate what I really mean (or at least I should mean if I got my philosophy straight) is that if I had decided to eat it, I would have: chocolate was available, and nothing and nobody would have prevented me from eating it. If the compatibilist gloss is correct, then determinism, when applied to mental processes and one's environment, is compatible with free will. It is true that given the states of my mind and my environment my eating vanilla ice-cream was inevitable, but determinism is perfectly compatible with the idea that had I decided to eat chocolate, I would have.

[2] See, for example, Earman, J., (1986), chapter III. See also Wilson, M., (1989).

Some philosophers, the *incompatibilists*, find the compatibilist position a rather cheap trick. For while determinism is compatible with the idea that had I decided to eat chocolate I would have eaten it, it is not compatible with the idea that I could have decided to eat chocolate. Saying that I would have done differently had I so decided merely moves the problem one step back unless I could have decided differently, and such capacity is forbidden by determinism. For the incompatibilist, one cannot be both a deterministic system and a being endowed with free will: one excludes the other.

Einstein, an incompatibilist who believed in universal determinism, concluded that we have no free will, and that therefore ultimately we are not morally responsible for our actions. He wrote to his friend Besso: "In relation to one's attitude in action, I agree with your remarks about loving one's enemies. However, for me, the intellectual basis for this view is the trust in unlimited causality. 'I cannot hate him because he *had* to do what he did.' So, I am closer to Spinoza that to the prophets. For this reason, for me there is not such a thing as 'Sin'." Analogously, in an essay honoring the Indian poet Tagore, he claimed that "causality does not stop before the majesty of our human will." [3]

Conversely, if one is an incompatibilist like Eccles or Stapp and believes that we do have free will, then one must find room for it in the natural world.[4] One might think

[3] Letter to M. Besso, January 6, 1948, in Speziali, P., (ed.) (1972): 392. A fragment of the essay in Tagore's honor is quoted in Home, D., (1997): 361. Spinoza, famous for his pantheism, belief in universal determinism and consequent lack of free will, was probably Einstein's favorite philosopher.

[4] Eccles, J., and Robinson, D. N., (1984): 98-100; Stapp, H., (2004), ch. 1.

that all one needs to do is to deny value determinism, an easy step given standard quantum mechanics. However, things are more complicated than that. If an action is ultimately the sole result of quantum indeterminacy, then it can hardly be attributed to the agent. Imagine a Geiger counter connected to the neurons charged with activating the motor nerves of your right arm. When an alpha-particle is emitted by a small quantity of radioactive material the counter clicks and causes the relevant neurons to fire, with the final result that your arm moves. It is hard to see how such movement would be free in the relevant sense because it would not stem from you: it would be as if you had a spasm from some neurological disease. The same would be true if the counter were biological, the radioactive emission in the brain, and its outcome the decision to move your arm rather than the moving of your arm: decisions, if one could call them so, would occur but would hardly be yours. Even within fixed boundaries and with fixed expected values, random events cannot provide free will on their own.

Still, if one is an incompatibilist convinced of being endowed with free will, invoking quantum indeterminacy may seem an appealing way to provide the physical possibility for free will, itself the result of the activities of a mind assumed not ruled by deterministic laws. The most radical of these projects involves the idea that the mind is a non-deterministic immaterial substance, a soul, capable of originating action and operating on the brain. We need not analyze the arguments put forth in defense of the idea that the mind is immaterial; typically, they revolve around the alleged inability of matter to give rise to consciousness or to explain its seeming evolutionary success. The fundamental problem facing interactionist dualism (the view that mind and matter are different kinds of substances and that they interact) has not been to explain how the immaterial mind could act on the material body, but how it could do so without breaking some conservation law or other. For example, Descartes, the originator of modern interactionist dualism, held that quantity of motion (akin to momentum) is conserved, and yet the immaterial mind, through free will, can spontaneously move the body by influencing the distribution of certain fluids (the animal spirits) in the pineal gland, a structure of the brain. Although we do not know precisely how Descartes reconciled the unavoidable introjection of momentum in the world caused by the mind's activity, there is some indirect evidence that he took quantity of motion to be the magnitude of momentum and that free will would only change its direction.[5]

How serious the problem was can be gauged by noting that Newton, a committed interactionist dualist, denied that momentum is universally conserved in part to assure the possibility of freely willed actions introjecting new force in the physical world. By contrast, Leibniz, who believed in the conservation of *vis viva* (twice the kinetic energy) thought that the soul and the body could not interact, and consequently adopted pre-established harmony, the view that the soul and the body have been so set up by God to develop independently but in coordination, like two clocks always marking the same time. Things got worse for dualist interactionism with the discovery of other conservation laws, such as the conservation of energy, and typically modern materialists appeal to conservation principles to discredit dualism.[6]

---

[5] The evidence on Descartes is Leibniz' word for it. For this and Leibniz' views, see Leibniz, G. W., (1705).

[6] For a contemporary example of the use of conservation laws against interactionist dualism, see Dennett, D., (1991): 35.

However, the advent of quantum mechanics and the widespread acceptance of the orthodox view have opened new areas of inquiry based on the idea that if the brain as a whole, or at least some significant parts of it, exhibit quantum mechanical behavior, then the mind might be able to control the relevant collapse events. If so, the mind would be able to choose one among different possible physical outcomes without impinging on any conservation law. The traditional physical objection against such schemes is that since the brain is large and hot, it is unlikely that as a whole it displays quantum behavior. It is true that some molecules with a diameter of about one nanometer (10^{-9}m) like carbon-70 molecules do display quantum behavior in double slit experiments up to temperatures of 1000K (the temperature of the brain is about 310K), and that the same has been observed in molecules double that diameter, suggesting that possibly some proteins may do the same. However, even assuming that some molecules and proteins engage in quantum behavior is not sufficient to introduce indeterminacy in the actual workings of the brain unless such behavior can be properly amplified to some significant brain structure, namely, one capable of exciting motor nerves. The problem is to discover the mechanism for such amplification. The difficulty is magnified by the very short times involved in decoherence. For example, Tegmark has argued that macroscopic coherence in the brain is lost in times of the order of $10^{-13}s$ to $10^{-20}s$, much too short for any psychological event to occur. To be sure, these figures have been challenged, lengthening coherence times by about 9 orders of magnitude, but even this does not seem sufficient, although it seems to place the available time somewhat in the order of what might be physically possible.[7]

[7] Tegmark, M., (2000). The rejoinder is in Hagan, S., Hameroff, S., Tuszynki, J., (2002).

One way to bypass the difficulty is to assume that the brain is a chaotic system, namely, a deterministic system highly sensitive to very small variations in the initial conditions, and that this sensitivity extends downwards, as it were, to quantum events. Then, the element of randomness present at the atomic level could be amplified at the neuronal level. However, not only there is no idea of how such amplification actually occurs, but there is no agreement on whether the brain is a chaotic system. The truth seems to be that we know too little about the brain's minute workings to establish whether this proposal is correct or even likely correct. A more promising avenue of enquiry is to note that although decoherence quickly eliminates interference terms it does not eliminate superposition, and consequently any approach not relying on interference terms is unaffected by it. One such approach has been put forth with a wealth of anatomical detail by John Eccles, a prominent neuroscientist and Nobel laureate, to whose views we now turn.

16.2 Eccles' Views

The neocortex, the part of the brain associated with higher mental activities, including decision making, is constituted by billions of neurons, the nerve cells making up the brain. A neuron is constituted by a body, by dendrites, fibers though which the neuron receives signals from other neurons, and by an axon, a fiber conveying signals to other neurons. The axon of a neuron ends in knoblike structures, the boutons, each of which is in close proximity (about 20 nanometers) to a dendrite or to the body of another neuron. An area in which two neurons are so connected is a synapse. The excitation of a neuron is governed by thousands of messages coming from other neurons, with each message being conveyed by exocytosis, the release of packets of about 5,000 to 10,000 neurotransmitter molecules, each composed of several atoms. These packets are contained in synaptic vesicles, themselves within a bouton, that release their content upon being penetrated by a few Ca^{2+} ions (Eccles, J. C., (1994)).

Although many more Ca^{2+} ions than are apparently needed to produce exocytosis are sent through the axon to each bouton, exocytosis occurs only about 25% of the times, a fact that Eccles has interpreted as a manifestation of quantum indeterminacy rather than as the result of an unknown classical mechanism. Together with physicist F. Beck, he has proposed a quantum mechanical model of exocytosis based on tunneling of Ca^{2+} ions into the vesicle's walls, interpreted as a quantum barrier. Each ion is represented by a wave-packet $|\Psi\rangle$ that upon impacting the potential barrier breaks into reflected and transmitted waves $|R\rangle$ and $|T\rangle$ so that $|\Psi\rangle = a|R\rangle + b|T\rangle$.[8] The superposition, in which each term represents a contribution to a possible action, is terminated by a collapse brought about by the intervention of the mind (taken by Eccles to be an immaterial substance, the soul) along the lines of Wigner's position (Ibidem, 154-59). By simultaneously producing the collapses of thousands of quantum states involving thousands of vesicles, we (our immaterial selves) control our brains, and therefore our bodies. This amazing influence of the self on the brain is the result of a lifetime of learning, witness the initially tentative and progressively more and more sophisticated attempts of infants to control the motions of their bodies (Ibidem, 172).

Obviously, if correct Eccles' theory does not impinge on any conservation law, thus making interactionist dualism physically possible, an achievement he repeatedly

[8] The reflected wave gives the probability that the particle will bounce back; the transmitted wave the probability that the particle will go though the barrier.

emphasizes, claiming to have 'transcended" the problem of free will, to wit of how the self can act on its brain given conservation laws (Ibidem, 173). However, the idea that I can actually bring about the collapse of hundreds of thousands of selected state vectors onto eigenvectors of my choice seems preposterous. To be sure, Eccles need not claim that I intend to produce the collapse of any state vector when I intend to move my arm any more than I intend to cause the appropriate sequential excitation of motor nerves when I intend to throw a ball. After all, I might not even know that motor nerves exist. Still, he needs to hold that I do control, at least unintentionally, the outcome of certain quantum collapses.

The problem is that nobody has ever managed to control the outcome of any quantum jump. Of course, one might deny the relevance of the objection. From the fact that I cannot move the ball on the table by telekinesis, it certainly does not follow that I cannot move *my* body; similarly, from the fact that I cannot control quantum outcomes in the laboratory, it does not follow that I cannot control them in *my* brain. However, in general there are physical means whereby I can move the ball on the table, but there are none permitting me to bring about the experimental outcome of my choice on a single superposed system. In short, objects I cannot move by telekinesis I can move otherwise, at least in principle, but the results of quantum collapse in the laboratory are totally beyond my reach; this makes Eccles' claim that I can control tunneling events in the synapses of the neurons of the neocortex rather suspicious.

A further problem is that in our experience collapse, even if mind-induced as Wigner claims, only occurs if a quantum system interacts with a macroscopic system.

However, no such system is present in Eccles'story.[9] Finally, there is no obvious evidence that the soul is not a deterministic system; after all, the laws of psychology, if they exist, may turn out to be completely deterministic. Still, if Eccles' account is even partially correct, he may claim to have shown how a non-deterministic immaterial system (the soul) could act on the brain compatibly with known physical laws, and this is no mean achievement.

A related proposal has been put forth by Stapp on the basis of the orthodox theory of measurement. In order to understand his views, we need to look at a rather strange result of quantum theory, the Zeno Quantum Effect.

---

[9] R. Penrose (who is not a dualist) has proposed a somewhat similar scheme. Having argued that human understanding cannot be computational (algorithmic, or rule-driven), that collapse is a physical event ultimately to be explained in terms of a (future) quantum-relativistic theory, and that such theory cannot be computational, he suggests that the interaction between the quantum and the classical (neuronal) levels in the brain manifests itself as consciousness. He locates the relevant quantum action at the cytoscheletal level (the cytoscheleton being a biologically defined structure that controls many of the cell's activities), and more precisely inside long filaments with the diameter of about 25*nm*, the microtubules that partially constitute the cytoscheletons of neurons. He suggests that the classical computer-like behavior of neurons is the amplification of cytoscheletal action, where freewill may be located (Penrose, R., (1994): ch. 7, especially 376-77). However, as he later notes, non-algorithmic process, such as consciousness and free will are in his view, can be fully deterministic. Hence, Penrose's position is somewhat peripheral to the present discussion.

16.3 The Zeno Quantum Effect

Zeno of Elea was a Greek philosopher who produced a set of paradoxes aiming to show that motion, and ultimately change itself, is illusory. The Zeno Quantum Effect (ZQE) is somewhat incongruously named after him. In standard quantum mechanics, it is a dramatic consequence of collapse, resulting in the stability of a system that, left to itself, would not be stable at all. In a way, this result is not surprising. If a system is in eigenstate $|\psi\rangle$ and an observation is carried out before the system has had time to evolve according to TDSE, then the system will remain in state $|\psi\rangle$. Hence, by carrying out a continuous set of observations on the same system, we can prevent it from changing even if left on its own it would change.

Consider an unstable system, such as a particle in an excited state. As long as time *t* is considerably shorter (but not too much shorter) than the natural lifetime for the transition to a lower energy level, the probability of observing the transition during time *t* is proportional to *t*:

$$\Pr(decay) = \alpha t, \tag{16.3.1}$$

so that

$$\Pr(no - decay) = 1 - \alpha t. \tag{16.3.2}$$

Suppose now that we have observed the system at *t* and we make another observation at time $2t$. Then the probability that there has been no decay is

$$\Pr(no - decay) = [1 - \alpha t]^2 \approx 1 - 2\alpha t \tag{16.3.3}$$

because *t* is very small. So, after two observations (16.3.3) gives the probability that the system is observed in its original state. Since as long as *t* is sufficiently short (16.3.1) remains true, if we had skipped the first observation at t and made only one observation at $2t$ we would have observed a transition with probability

$$\Pr(no-decay)=1-2\alpha t, \quad (16.3.4)$$

which is the same as (16.3.3). In brief, the first observation made no difference. However, if time t is *very* short (how short depends on the system at hand), the probability of observing the transition during t becomes

$$\Pr(decay)=\alpha t^2, \quad (16.3.5)$$

so that

$$\Pr(no-decay)=1-\alpha t^2. \quad (16.3.6)$$

Hence, after two observations, one at t and one at $2t$,

$$\Pr(no-decay)=\left[1-\alpha t^2\right]^2 \approx 1-2\alpha t^2. \quad (16.3.7)$$ [10]

If we had skipped the first observation at t and made only the observation at $2t$, we would have observed the transition with probability

$$\Pr(no-decay)=1-4\alpha t^2, \quad (16.3.8)$$

which is *smaller* than (16.3.7). Consequently, when t is extremely short, observation tends to decrease the probability of transition, and therefore tends to keep the system in its initial state.

Suppose now that we divide a time T into n extremely short parts and make n observations, one for each part of T. Then, at T, after n observations

[10] The value of t for α-decay for an atom is of the order of $10^{-21}s$, a phenomenally small time.

$$\Pr(no - decay) = \left[1 - \alpha\left(\frac{T}{n}\right)^2\right]^2 \approx 1 - \frac{\alpha}{n}T^2, \qquad (16.3.9)$$

which becomes closer and closer to 1 as *n* becomes larger and larger. If in the limit *n* becomes infinite, that is, if the system is under continuous observation, then it will not decay. Experiment seems to confirm that ZQE is real, although not everyone is convinced.[11] We can now turn to Stapp's views.

16.4 Stapp's Theory

For Stapp, as for orthodox quantum theory, measurement involves three basic steps. [12] The researcher decides what experiment to perform by setting up the appropriate measuring device, thereby determining the basis in which the state vector will be expanded; TDSE governs the entanglement of the system and the measuring apparatus; finally, collapse determines what measurement return one gets. There is a choice determined by the experimenter's evaluation of what to do, a deterministic development ruled by TDSE, and a probabilistic event expressing Nature's 'answer' to the question posed by the experimenter. As the measuring device is made up of atoms, ideally we can describe it quantum mechanically in terms of a reduced density operator obtained by tracing over the density operator of the rest of the universe. The experimenter can then be thought of as a mind (a stream of consciousness) whose mental states are in parallel with the physical states of the brain, and the brain as the part of the physical world to which the mind is directly related.

---

[11] For a discussion, see Whitaker, A., (1996): 308-314.

[12] Stapp has presented his views in numerous papers. See, for example, Stapp, H., (1998); Stapp, H., (2004a); Stapp, H., (2004b).

Like Eccles, Stapp focuses on exocytosis. The calcium ions responsible for the release of the transmitters in the vesicles emerge from ion-channels about 1 nanometer wide. Since this allows a rather close determination of their positions, Heisenberg's Uncertainty Principle generates a corresponding uncertainty in their lateral velocities. As they must travel about 50 nanometers to reach the vesicles, the lateral spreading out of the wave-packets is about the size of the calcium ions itself, thus generating a situation similar to the double slit experiment. The result is a superposition at the (macroscopic) brain level analogous to that proposed by Eccles, in which each element, expressing the entanglement of a great number of items, represents a possible template for action, a possible pattern of neuron firing leading to an action such as the lifting of one's finger.

Now Stapp assumes that as in a laboratory experiment the researcher can choose the basis in which the state vector is expanded, so the mind can choose, presumably non-intentionally, the basis in which the reduction of the state vector will take place. As in the case of the researcher, quantum mechanics is silent on how the choice came about, subsuming it under the rubric of free will, by which Stapp means something that is not subject to determinism and in addition is not controlled by the known laws of physics, or, which is the same, always treated as an independent variable.[13] Collapse will then occur according to quantum rules (contrary to Eccles) and the corresponding mental experience E (for example, the experience of intending to lift one's finger) will take place. At this point, on the basis of evaluative considerations about which science is at least for the moment silent, if the mind approves of E, by 'choosing' the same basis as before it can very quickly bring about a new collapse *before* the brain has had time to evolve

[13] Time in Newtonian physics is always treated as an independent variable, for example.

significantly, thus activating ZQE. The result is the familiar experience of willful effort whereby I keep intending to lift my finger as I actually lift it. The mind's choice of basis and its approval of E is where free choice, that is a choice not governed by known physical laws, occurs.

As we noted above, Eccles' and Stapp's proposals are similar: both focus on exocytosis; both believe that classical physics cannot provide any account of consciousness, with the result that if classical physics is taken to be causally closed, then consciousness becomes at best a useless appendage without any causal efficacy. There are, however, two significant differences between them, one physical and the other philosophical. The physical difference is that while Eccles resorts to the stratagem of endowing the mind with the capacity of altering expectation values in exocytosis, Stapp takes the more cautious view that the mind's influence on exocytosis occurs in accordance with a relatively known quantum mechanical phenomenon, ZQE, thus avoiding the problems associated with Eccles' gambit.

The philosophical difference is more nuanced. Both Eccles and Stapp believe that using classical physics to give an account of the workings of the brain is both erroneous and anachronistic; however, while Eccles is an unrepentant dualist, Stapp's position is less sharply defined. He does believe that contemporary quantum theory involves some sort of interactive dualism, but he also seems to lean towards a form of dual-aspect monism in which the world is made of only one type of stuff that manifests itself as mind or matter. Quantum physics shows that physical reality behaves more like some sort of spatially encoded information governing the tendencies for experiences to occur than like any sort of traditionally conceived matter. Stapp's position, however, is not exempt from difficulties even if one adopts the orthodox interpretation. For one thing, the frequency of mind-generated choices necessary to activate ZQE is phenomenally high, making it doubtful that it is in the range of mental activity.