Free Will

Welcome to Physics and Religion, a series of podcasts exploring the interaction between modern science and traditional religion. I am your host, Dr. A. S. We have been trying to understand consciousness, but neither has the neuroscience of vision nor have the claims of artificial intelligence nor has physicalism with all its objections given us much understanding. Let's try another approach and look into the phenomenon of free will.

On an early morning in May 1987, Kenneth Parks got into his car and drove 23 km. to the house of his wife's parents. He attacked both of them with a kitchen knife, killing the mother and leaving the father seriously injured. He awoke to find himself covered with blood. Terrified he drove to the nearest police station and turned himself in. At trial, Parks argued that he was automatistic and not criminally liable. From the doctor's evidence it was determined that the accused was sleepwalking at the time of the incident, and that he was suffering from a disorder of sleep rather than neurological, psychiatric or other illness. Five neurological experts also confirmed that he was sleepwalking during the time of the incident. The jury acquitted Parks. The case went all the way to the Canadian Supreme Court. Parks was acquitted at each step along the way.

There are similar cases on record from other countries, but this one is especially clear cut and puzzling. If I killed my in-laws I would be convicted. It would be assumed that I was aware of the difference between right and wrong, that I was free to choose between them, and I in fact chose the wrong. Parks on the other hand did not have free will. He could not choose; he was not guilty. Here our common sense hits a snag. First, was Parks conscious? He was able to get out of bed, get dressed, and drive his car to the distant location. No doubt along the way he obeyed all the traffic laws and stopped at all the red lights. He found the house and found the kitchen knife. If we assume that he was not conscious, then he calls to mind our old friends the philosophical zombies. He behaved in all respects like a normal person but he was not conscious. Why did he commit the murder and attempted murder? Again the facts of the case are particularly clear cut. The evidence presented in court showed that he was on the most cordial terms with

his in-laws. Where then did the motivation to commit murder come from? Finally, if he did not have free will when he was sleepwalking and did when he was awake, it seems that free will is something that can be switched on and off. If so, where is the switch?

Our intuitions as well as our legal system are convinced that we have free will. There is a logical argument to that effect as well. If I did not have free will, I would be under no compunction to tell the truth. What I am telling you might be complete nonsense or it might be an assortment of lies. If you are to take seriously anything I say you must assume that I have free will. Unless we make this assumption for everyone, communication and community become impossible. With this in mind you may be surprised to know that among philosophers and neuroscientists, free will is the minority opinion. Why might we not have free will?

There is a physics argument that goes back to Einstein. I need to tell you one simple thing about relativity; simultaneity is relative. What that means is simply this; from our perspective, Thursday comes after Wednesday. In principle there could be observers moving relative to us a speed close to the speed of light for whom Wednesday comes after Thursday. The fact that this is wildly impossible in practice is of no consequence; in principle, it is a simple consequence of relativity. If you ask, what day *really* comes first, Einstein would reply that this is a meaningless question. Simultaneity, and with it the whole business of what comes first, is relative. Now let me tell you a story about the Vogons. If you have read *The Hitchhiker's Guide to the Galaxy* or seen any of the various movie adaptations, you will know that the Vogons are an intensely bureaucratic people. Things can only be decided after many meetings and endless filling out of legal forms. In this spirit there was a meeting held on a Wednesday to decide whether to destroy the earth to make way for a new hyperspace bypass. The decision was positive and the war ships were launched the following day. As you know, the earth was destroyed many years later. The point is this, for some observers the ships were launched before the Vogons decided to launch the ships, and of course, the Vogons would never allow such a thing. How are we to understand this? Einstein's answer is that past, present, and future are in place. Time is not

like a river that flows past us. Rather we move through time encountering all our life's events as we go. I can't even say that the past, present, and future exist now or that they have always existed, because "now" and "always" carry some sense of temporality. Back to the Vogons: Imagine a landscape where there are two monuments, one labeled "Decided" and one marked "Launched." People are wandering through this landscape. Some will pass "Decided" and then go on to "Launched." Others will pass "Launched" and then go on to "Decided." Physicists call this "block time." It implies that everything you will do in your life is already in place. It seems that you have no choice; but the more I think about it the more ambiguous it becomes. Tomorrow, do I dare to eat a peach? (To quote T. S. Eliot) My decision is already in place in the landscape of events, but perhaps my freely deciding to eat or not to eat the legendary peach is also part of the landscape. This paradox illustrates the difficulty we have in explaining exactly what free will is; but it also calls to mind the principle I enunciated in connection with solipsism. A theory that cannot be proved wrong under any circumstances should not be taken seriously.

I would like you to think about the following argument. Physicists like what we call operational definitions. We think a concept is not meaningful unless you can propose some measurement or experiment to define it. Whether the experiment is possible in practice is of no consequence. For example, the unit of charge is the coulomb. It is defined by saying that two charges of one coulomb each placed one meter apart would exert a certain (enormous) force on one another. (The experiment is not possible. One coulomb is a huge charge. The force is inconceivably large.) Can you think of an experiment, however impracticable, to decide once and for all whether anyone has free will. My idea is to use a time machine. Our test subject, call him Albert, makes a decision between two alternatives, A and B. He chooses A. We send him back in time to yesterday. If this test is to be meaningful, we must set everything back one day, *including his memories*. Now we will see the second time around whether he chooses A or B. Alas – If he chooses A, we could argue that he chose A because that was in fact the best choice and after free consideration he chose it again. We could also say that he chose A because he had no choice. If he chooses B, we could argue that

that proved that he was free to make a different choice this time. We could also assume that there was some random unpredictable element to his choice and he randomly chose B. So far as we can tell from this thought experiment, free will is a meaningless concept. I will return to the issue of meaningless in the next podcast. For the time being, let us see what the philosophers have to say.

Here is the classic argument. Everything that happens in the universe happens because of some prior chain of events. Our thoughts and actions are events in the universe. Therefore, they are caused by some prior chain of events, which may go back to the beginning of the universe. Therefore, we have no free will. A more modern formulation would have to allow for the fact that quantum mechanical events happen randomly and unpredictably and complex nonlinear systems, which include almost everything in the universe, tend to behave in completely unpredictable ways. It is not clear that this saves free will, however. It rather means that our behavior, rather than being predetermined, is mostly random for reasons over which we have no control. According to Wikipedia, “The field remains highly controversial. There is no consensus among researchers about the significance of findings, their meaning, or what conclusions may be drawn. The precise role of consciousness in decision making therefore remains unclear.”

But is this really incompatible with free will? At this point,… the philosophers divide into two camps called, not surprisingly, compatibilists and incompatibilists. The compatibilists claim there is a flaw in the argument I just presented; freedom and the sort of determinism I have just described are not inconsistent. If we define carefully what we mean by freedom we will see that we can have our cake and eat it too. Freedom they say is the ability to do what we want to do, to choose what we want to choose. Suppose I am given a choice between alternatives A and B. If I chose A because I wanted to and could have chosen B if I had wanted to, then I am free. If I am coerced to take A then I am not free. There are many situations, however, in which this doesn’t make sense. A drug addict for example, desperately wants a fix. He has the money and his favorite dealer is ready to make a sale. It seems odd to say that the addict is free when he is driven by his addiction. Hypnosis makes another good counterexample. The hypnotist

can say, “When you wake up you will badly want to give me your money.” The subject awakens and complies.

Another try at compatibilism goes like this: you are free to make a particular choice if you could have done otherwise. I offer a bizarre thought experiment that suggests otherwise. Suppose you are given the usual two options A and B. You want to choose A and you eventually do, but unbeknownst to you our friendly hypnotist has implanted a posthypnotic suggestion; if you should ever change your mind and decide to choose B a hypnotic trigger will force you to choose A. So you freely chose A even though you could not have done otherwise. This is a way of introducing the idea of second-order desires. A second-order desire is a desire about desires. Think about the drug addict. He desperately wants his fix but he wishes he didn’t have such desires. This is a second-order desire. The addict is not free because he cannot act in accordance with his second-order desire. But even this doesn’t capture the essence of freedom. Suppose our drug addict is so habituated in his addiction that he values his lifestyle. He looks forward to his daily trips to his dealer. Every morning he wakes up looking forward to his day, “Ah,” he says, “Today I’m really going to get high.” He is acting in accord with his second-order desires but he does not really seem to be free.

Our final try at compatibilism is called contextualism. Perhaps the meaning of freedom depends on context. Here is an example. Suppose you are robbed at gunpoint. Our clichéd mugger says, “Your money or your life.” Perhaps in the context of the nexus of the universe you are not free. Who cares? Leaving that aside, are you free to refuse? Yes and no. Yes if you don’t mind having your head blown off. No if you want to make a wise decision regarding your future life. Perhaps the word “freedom” is meaningless without some specification of context. This brings us back close to common sense. Perhaps every complete specification of context should contain clauses such as, ⋯in the absence of drug addiction, ⋯in the absence of armed robbers, ⋯in the absence of federal or state laws forbidding the action, ⋯I the absence of mad hypnotists, etc. When we judge a man guilty or innocent of some crime it is always within some implied context. Was the accused sane and mentally competent? Are the relevant laws

applicable? Was the evidence handled properly? The laws of classical and quantum mechanics are simply not relevant in this context.

Welcome to Physics and Religion, a series of podcasts exploring the interaction between modern science and traditional religion. I am your host, Dr. A. S. We have been struggling to understand free will. In fact, we have been trying to figure out what it is in the first place. I suppose our arguments could be characterized as attempts to define free will in such a way that we have it. What does science have to say? There was a classic experiment performed by Benjamin Libit and his colleagues at the UC Medical School in San Francisco during the 1980's. Volunteers were connected to an electroencephalograph (EEG for short) and asked to make some small trivial movement such as tapping a finger at random times of the subject's own choosing. It had been observed that under these circumstances an odd electrical signal called the readiness potential appeared in the electroencephalogram. Libit then asked his volunteers to make note of the time at which they decided to move their finger. He discovered that the readiness potential began to build up 350 ms before the subject decided to act. The implication is that the decision was made unconsciously. The subject was only under the later illusion that he or she had made the decision to act consciously and deliberately. Therefore, it was concluded, there is no free will.

This experiment and its interpretation have been criticized, defended, reinterpreted, revised and replicated. The experimental data are easy to summarize, however. The experiments have all involved on-the-spot decisions to perform trivial tasks. In these cases, some sort of neuronal activity preceded the subject's awareness of having made a decision; the rest is interpretation. I would like to pass along two interpretations; you are invited to invent your own. First, the readiness potential may be what the computer science folks call a gating

signal. In itself it does not make the decision, but it must be present for the decision to be made. We say that it “gates” the decision-making circuits.

My second suggestion requires more explanation. Perhaps there is some subsystem in the brain that is thinking rationally, at least some of the time, and is making decisions freely, again at least some of the time, but we are not immediately aware of what is going on there. I have two analogies that might be relevant and helpful. Modern surgeons often operate with microscopic instruments that work by remote control. The surgeon might be in another room viewing the surgical site on his computer monitor. The surgery is happening inside the patient; the surgeon has a remote computer’s-eye view of what is going on. Perhaps our consciousness is like the computer monitor. The real action is taking place somewhere else, but that does not necessarily mean that the thinking isn’t being carried on competently and freely. My second analogy concerns blindsight patients who can perform simple tasks in their blind field. One patient was reportedly able to play a creditable game of darts even though he could not “see” the darts or the board. Apparently, some subsystem in his brain was processing visual data and relaying it to his motor centers even though he was not conscious of seeing anything. The point is that there can be parallel subsystems in the brain that process similar data even though we may not be conscious of both of them.

Libit’s experiment as well as similar experiments that have been performed subsequently all involve trivial, meaningless decisions that are made for no particular reason at no particular time. Claiming that this precludes the possibility of free will in general is a monstrous over interpretation.

I would like to add a coda to this entire series by dropping back for once squarely in my own field of expertise, quantum mechanics. There was a famous debate between Albert Einstein and Neils Bohr over the nature of reality. I am going to explain it with a specific example, but the question pervades all of quantum mechanics. Electrons have a property called spin. An ordinary sphere makes a good analogy. It can spin around its axis at any speed. We say that the direction of the spin is the direction that the spin axis points. The speed with which it is spinning is proportion to something called spin angular momentum or “spin” for

short. Thus every spinning object has some spin and the spin has a direction. The same is true of electrons but in an odd way; electrons are point particles, they have no size, how can they spin? The answer is that any spinning charged sphere would generate a very specific kind of magnetic field and electrons have exactly that magnetic field. The second odd thing about electron spin is that every electron in the universe has the same spin! The oddest thing about the electron is this: whenever you try to measure the direction of the spin, you always get one of two answers: up or down. Up or down with respect to what you ask. The answer is dumbfounding, up or down with respect to the measuring device. No matter how you hold your measuring device, the answer is always up or down. You can make 100 successive measurements in 100 random directions, and the answer will always be either up or down. How can this be? Here is where Bohr and Einstein had their famous disagreement. Einstein thought that the electron spin always had a direction, we just don't have the theory to understand this; QM was wrong or at least incomplete. Bohr thought the question was meaningless. We create the direction by measuring it. Time has proved Bohr more nearly right than Einstein but the issue is still being debated.

Let's assume that Bohr was right, then QM has forced on us a profound concept. Some apparently well-posed questions cannot be answered because they refer to a reality that is permanently veiled. Not as Einstein thought, veiled until we get the right theory, but for some perhaps metaphysical reason, permanently veiled. Not, as others might claim, because we are not smart enough, but because reality is so constituted that these things are not knowable. I'm sure you can see where I am going with this. Both in our investigation of consciousness and of free will we were frustrated. What is consciousness? Do we have free will? Not only did we have trouble answering these questions, we had difficulty understanding what the questions meant. The barrier to our understanding ultimately was the disparity between the first- and third-person perspectives. Perhaps these two realms are separated by a veil that is every bit as impenetrable as the one that conceals the electron's spin. Perhaps this entire series has been about what Bohr would call meaningless questions. I take this as good news. QM is a vibrant theory

firmly rooted in the meaninglessness of meaningless questions. Perhaps this is a good model for life in general.

A similar situation occurs in classical chaos theory, although it is not widely acknowledged. For the purposes of illustration, consider a simple nonlinear system; perhaps two pendulums connected by a spring or one pendulum hung from the bottom of another. These systems are subject to Newton's laws of motion as elaborated by some of the great mathematicians of the 18th and 19th centuries. The result of their formalism is a set of coupled, nonlinear partial differential equations. To this extent, the systems are deterministic, but at this point, a startling problem appears. The equations typically have no solutions! It is not the case that we are not smart enough to solve them or we lack the necessary mathematical machinery – they just don't have solutions. What does this mean you ask; if we constructed one of these systems and set it in motion, it would do *something.* Yes, but what it would do is unpredictable. You might set it in motion an arbitrarily large number of times and it would never do exactly the same thing twice.

There are powerful computer-driven algorithms for obtaining approximate solutions to equations like these. With the help of programs like these, we can see what is really going on. In some cases the system's behavior becomes wildly chaotic. In other cases the system seems to be executing an elaborate dance which proceeds according to a precise set of rules but in which no one step is ever repeated. As the doctor once said of my heart beat, it is regularly irregular.

What can we learn from these simple systems about the universe as a whole? First, almost everything is the universe is nonlinear. The sort of simple problems that have exact solutions appear in undergraduate textbooks but nowhere else. In some vague qualitative sense, what I said about double pendulums is true of all mechanical systems. Now recall the philosopher's proof that we have no free will. Every action stands at the end of a long chain of cause and effect that might go back to the beginning of the universe. To this extent, everything we do is predetermined. There might be exceptions to this if quantum mechanics were involved at some point, but then our actions would be wildly unpredictable. Now

we can see the truth of the matter. The chain of cause and effect is not clear cut. It becomes increasingly fuzzy as we try to trace it back into the past. Nonetheless, our behavior need not be completely unpredictable. There is a window of opportunity between complete determinism and complete chaos. We are caught up in a dance in which there is a definite order but in which each step is new. In this way it seems to me, physics leaves a gap in which free will can operate.