D.D. – Daniel Dennett

D.V. – Dmitry Volkov

V.V. – Vadim Vasilyev

V.V.: Our first question. You often claim that there is no hard problem of consciousness at all, just a vast amount of easy problems. But what about the mind-body problem? Does it exist? And if it does, what is it? And is it possible to solve this problem?

D.D.: I think the solution to the mind-body problem comes from first taking a deep breath and trying to adapt first the scientific perspective where we consider ourselves as animals: mammals, primates moving around on the planet. And in fact you might actually take a trick from Descartes here who wrote about *le monde,* saying "Imagine a planet," and after he described it, he announced 'Oh! That's our planet!' So we have to "imagine going to a planet and finding things that look like us," as if we were coming here from the outside. And the question arises: is there the mind-body problem? Do we discover the mind-body problem on this planet? And of course we do. The problem arises because we've got to interpret the beings that we discover, we've got to treat them heterophenomenologically, we've got to treat them from the intentional stance because there seem to be agents there fending for themselves very well. So we start a process of interpretation.

First of all we do the physiology and the physics and so forth. But then we still have this project of interpretation. And what we focus on, if they really are like us, is their behaviors, including the fact that they are apparently communicating. We learn their language. So now we can really use heterophenomenology; we can ask them, 'Tell us what it's like to be you.' And they do! Now suppose the hard problem arises when you think "Okay. Well at least there are zombies present—they talk, they are intelligent. But are they like us? In what regard? Is there something that it is like to be them? Do they have an inner life?" But of course they do. Listen to them! There's no doubt about that.

It recently occurred to me that there is a very simple way of showing how strange the confusion about this matter is. It's to think about how more than fifty years ago, Wilder Penfield in Montreal was doing brain surgery on wide-awake patients. He took the tops of their skulls off, their scalp and so forth. They were sitting in a chair like a dentist's chair and they were wide awake with the top of their scalp off, and he touched points on their cortex with electrodes and asked them to talk about what was going on. And they said, "All that I experience is a tune, or there is my finger twitching," and so forth. I've actually shown the film Penfield made of this to some of my students, and of course thousands of people have heard about it. I've never heard anybody say " Oh my God! There is nobody home in there. It is just a zombie. Look! They take off the top of the scalp, and It is just a brain, that's all that is in there!" It's too obvious that the people are conscious. And it would be just as obvious if we took the scalp off and we found a machine bristling with transistors inside. I mean there is a huge leap of imagination required to understand how a conscious life can be implemented in either silicon or in proteins. But it can.

Penfield's results thus give a kind of immediate proof of this point. We know that there's some kind of brain machinery that does the job, and the further idea that there isn't a mind present in these cases at all is, I think, actually incoherent. After all, suppose there are zombies. Then to try to imagine a zombie, you have to remember that by mutual agreement the philosophical definition of a zombie is somebody who is behaviorally indistinguishable from a human being. In other words he is good company, he gets jokes and has complex preferences. You can engage such a zombie in indefinitely iterated, recursive self-reflection.

I recently wrote a bit which will come out in my next book [INTUITION PUMPS] in which I invite readers who really want to imagine a zombie to read a novel by Jane Austen or by Vladimir Nabokov or by J. D. Salinger. I ask them to try to imagine that the novel is actually a story about a zombie. I made a little discovery when I tried to do this myself. As you know, novelists have different perspectives that they can adapt. There is the omniscient author who can go into the minds of all the characters and say, for example, "then she wondered about the weather" and so forth. And then there is first-person narration, in a novel like *Moby Dick* in which the whole story is told through the mouth of one character, who begins the novel: "Call me Ishmael." In *The Catcher in the Rye* by J.D. Salinger, Holden Caulfield is a troubled teenager talking to you. I point out that, curiously, a first-person narrative novel is easier to imagine as a zombie novel then is a novel with an omniscient author, because in the case of the first-person narrator, you are just seeing the behavior, you are just hearing the voice. Whereas the omniscient author pretends to be reading the minds of all the characters. And in fact of course the case of the omniscient author reflects a true situation—we do read people's minds (in our imagination) all the time. What the novelist who employs omniscient narration does is only an artful extension of what we all do. Namely, we imagine what's going on in other people's heads. We can test hypotheses about what those people are thinking and experiencing as exhaustively as we choose—that's what heterophenomenology does. Once we've got the science for how the brain makes each bit of that understandable, there is no more problem.

I had a very interesting discussion with Tom Nagel about this topic around the time that he wrote his essay 'What is It Like to Be a Bat?' He said, "Well, neuroscience could give you a perfect correlation between brain events and conscious events, but it wouldn't explain the correlation." I said, "Okay. Let's explore that. Can you give me an example in which we have more than just correlation?' And he said, "Yes. Chemists and physicists can tell you why water is wet. The macro-property of wetness can be explained—not just correlated with, but explained in terms of—the bounce of water molecules, and so forth." I said, "Okay. That's a nice example of explanatory correlation. And how about when industrial chemists create a new polymer that's never been seen before, an artificial polymer? Before they have even made it, they tell you what its tensile strength will be, what its color is, whether it's going to be brittle or stretchy and so forth. They can do that sometimes." He said, "Yes." I then said to him that in that case we've already entered the age of explanatory neuroscience. I showed him the Ramachandran and Gregory motion-capture experiment in which they predicted a visual illusion that had never been seen before. They predicted it on the basis of their knowledge of the organization of the visual cortex in terms of the blobs and the interblobs, and the parts that are motion-sensitive and the parts that have poor spatial resolution. I said, "Isn't that exactly analogous to what the industrial chemists do? These folks knew enough about how the brain works to predict novel experiences in advance. So they can say, 'We will prove that we know how the brain works, because we will put together the circumstance which creates an illusion never before seen, predicted in advance.'" That's why I think there is no hard problem.

D.V.: In *Sweet Dreams* you said that reverberation or the echo system is what constitutes the core feature of consciousness. The book was written seven years ago. So in recent studies on animals or on child development, have you seen any information or any empirical facts that would provide evidence in support of this hypothesis? And do we know now which animals have the echo systems?

D.D.: There may well be such evidence, but I haven't particularly had my attention drawn to any. I think, however, that we are seeing a rapid development in research. I wish I had been able to put the theory in my book in terms of the new developments in Bayesian predictive coding in the nervous system. That seems to me very clearly to be an implementation of the echoing system that I described, except for one

wonderful detail. That detail is that you might say that this system is a “null echolng system” because the Bayesian predictive coding works like this: when the brain anticipates a situation, makes a prediction, and enough of the prediction is confirmed, then no signal comes back at all. A signal only comes back for disconfirmation. So the signal gets tuned by selective disconfirmation, but as long as the prediction turns out to be true, there is no echo.

I have an example just to give people the flavor of this how this works. Suppose I want to know how things are back home. I’m on a trip, and I arrange to call up my wife. Here is what I do. I just say, “Dear, I'm going to talk about everything in our house and the people we know. And don't say a word unless I say something false.” I talk for ten minutes, and she doesn't say anything. I can conclude: “Oh! Great! Everything is fine. Goodbye.” Then I hang up. That's the underlying idea of predictive coding. It seems to be a very powerful technique. And it seems to explain features of neuroanatomy which are otherwise quite baffling, such as the fact that there is more out-bound traffic in the visual system than there is in-bound traffic. There are more downward connections going from central areas back down to the eye than vice versa, which at first seems strange. But I think that that inversion is really very important. I'll be talking about that later today.

V.V.: Let us clarify one point of your theory when you say that consciousness in some sense doesn’t exist. Are you ready to admit that it, at least, might exist? Is it possible according to you to conceive the existence of consciousness in the very sense in which you denied its actual existence? And if it is possible at least to conceive the existence of consciousness in that sense, then in what respect does that conceived situation differ from the actual one in which, as you hold, consciousness doesn’t exist? What should be added to the actual state of the world in order to arrive at a situation in which consciousness does really exist? Or is it your view that the existence of consciousness is inconceivable in the same way that a round square is inconceivable?

D.D.: Good question. I think philosophers have made a mess of thinking about conceivability and imaginability. They say, ‘”Well, I can conceive that zombies (for instance) are possible” or “I find that such-and-such is inconceivable.” But conceiving of something is hard work, and you seldom can be sure when you have succeeded at it. The history of science is full of cases of people who said, “That is just inconceivable.” And then later anybody could conceive it. I have a lovely passage from the great geneticist William Bateson in 1916.

> The properties of living things are in some way attached to a material basis, perhaps in some special degree to nuclear chromatin; and yet it is *inconceivable* [my emphasis—DCD] that particles of chromatin or any other substance, however complex, can possess those powers which must be assigned to our factors or gens [= genes]. The supposition that particles of chromatin, indistinguishable from each other and indeed almost homogeneous under any known test, can by their material nature confer all the properties of life surpasses the range of even the most convinced materialism.

Bateson simply could not imagine a three-billion-codon-long double helix of DNA packed into every cell, but today children learn this in elementary school. So we have to be careful about what is inconceivable. And what is conceivable, for that matter. I think the question, “Can I conceive of consciousness being something extra, something added to all the information-processing occurring in the brain?” is deeply misguided. In one sense, sure, I can! Nothing easier! I can also conceive of massless gremlins that live inside the pistons of every internal combustion engine. They are basically undetectable by any known physical test, but—I insist I can conceive--they are in there. So what? This cheap sort of conceivability doesn’t create a Hard Problem for internal combustion engines.

V.V.: So the situation is the same with consciousness? Is that what you want to say to us?

D.D.: Yes, I would say it's the same. I don't think any automotive engineer or physicist would worry about the fact that they can't rule out invisible gremlins. I don't think neuroscientists or cognitive scientists should worry about the fact that they can't rule out this extra consciousness. It's a trivial possibility.

V.V.: But what is this extra consciousness like? How do you conceive it, this extra consciousness which actually doesn't exist, but which can be conceived?

D.D.: That's a sort of "mug's game," as they would say in England. It's a game with no rules. I think that, although they might not admit it, when a lot of people think of consciousness in this extra sense, they simply imagine a sort of radioactive glow, a special light that is either turned on or not.

V.V.: Really? Do you believe that?

D.D.: I think that the way people think of such an extra consciousness really is something like that. It's not radioactivity, of course, and it's not warm. Some of them imagine it as arising whenever there is some kind of critical mass. When the conditions are just right: SHAZAM!—this extra glow just happens. {<= "shazam" is an expression in colloquial English that was introduced by conjurors in the 1940s to indicate that an extraordinary event or transformation is about to happen or has just happened: "snap your fingers, and—shazam!—you'll be in China"} I agree that if the idea of an extra consciousness is articulated like that, it looks quite silly. So of course they don't articulate it that way. But they don't articulate it in *any* clear way, and sometimes they acknowledge that they can't, and make a joke about it, as when Ned Block quotes Louis Armstrong on what jazz is: "If you gotta ask, you ain't never gonna know!" Until they articulate a positive account of what more is added by the special sort of consciousness they think we have and a zombie lacks, they don't give us anything to talk seriously about. Let them articulate, clearly, what they think this extra thing is, and then we can discuss it.

V.V.: Feelings, emotions, mental images, these are all conscious states. They are not gremlins! Why think of them as being gremlins? Emotions are examples of consciousness—joy, and so on.

D.D.: Let's do emotions.... Christoph Koch is a romantic about this, and he is a die-hard believer in pains and emotions over and above the functional (and dysfunctional!) properties they exhibit. He once wrote me a letter about a toothache that he had when he was climbing in the mountains. "You tell me there is no toothache." No, I replied, I am saying there is no *extra quale* of pain over and above the effects . I asked him to Imagine two treatments, and then tell me which one he wanted. In treatment A, he is no longer distracted, he can think about anything he wants to, he is not obsessed with anything about his teeth; he is cheerful, and he can conduct his life without any interference. But (I said) there is nevertheless intense pain in his tooth all the time (whatever you think that means). In treatment B, *that* pain (whatever it is that treatment A doesn't eradicate) is completely gone from his tooth; but he can't stop thinking about his tooth, he can't read, he can't enjoy food, he can't make love, he can't have any pleasure in life at all, because the damn tooth keeps drawing attention to itself. But, still, there is no pain (of that awful but indefinable kind that treatment A fails to treat). Koch said he would choose treatment B, which makes me admire his devotion to consistency, but wonder about his powers of imagination. From this reply to Koch, you can see how I will deal with this kind of case in general. When people suppose that the pain is in some way really *there* as a separate thing, there is something as simple as an error theory of subtraction going on. Suppose you say that you have an image of pain, for example, your folk image of pain. Then I will start subtracting things. We take away the distraction, we

take away the negative effect on performance, we take away the interference with pleasurable activities. We take away, we take away ... And people think, “Okay, but you haven't taken away the pain yet.” But how do you know? I think that, in this case, a number of little problems constitute the apparent big problem, and if you remove the little problems, no big problem remains. But of course your website talks about the hard problem. You are taking it seriously.

V.V.: In our previous interview you said, by the way, that you cannot conceive that I am a zombie as regards my mental imagery. To explain your point, you provided a lovely example by asking me to combine the letters D and J into the shape of an umbrella. I managed to do that, and so you were sure I was not zombie. I had mental imagery. You were sure that I couldn’t have succeeded in combining the letters without using mental imagery. However, after the interview you said that these images are better treated as structures not as pictures. So it seems that you can conceive that I have no mental pictures. But, as a matter of fact, I have such mental pictures. So I conclude that you can, after all, conceive that I am a zombie if by zombies and mental imagery we understand subjects without mental pictures. Is this line of argument right or not? Where is my mistake?

D.D.: Well, I have an extended example in *Consciousness Explained* that discusses this kind of point. The example is Shakey, the early robot. And we see on the external screen the mental images Shakey is manipulating! But, in this case, if you turn off the screen, Shakey’s performance is not affected; Shakey is not looking at that screen. What’s inside the computer? Well, there is no image in 3D space. The “image” is just an array of zeroes and ones, as you know, and these individual pixel-representers can be stored anywhere in the computer’s memory. So, first of all, let’s get clear what you mean by image. In an old fashioned film, cinematography film, there are lots of images; each frame has an image. And in the case of a digital video, are there images? If you have a DVD, are there images on the DVD? Or just instructions for making images?

V.V.: Instructions, of course.

D.D.: Aha! Maybe all we need is instructions for making images, and maybe that what's manipulated in your brain. How could you tell that you had actual images or that you had just instructions for making images which you didn't follow by making an image but just manipulated?

V.V.: But I see a difference between structures and pictures. And you admit that you accept this difference. I agree with you that there is difference between structures and pictures. But I not only have structures, I also have pictures.

D.D.: No, you think that you have pictures.

V.V.: Maybe. That’s my mistake? It just seems to me that I have pictures?

D.D.: It's just seems to you that you have pictures. We can test that claim in a way. Steve Kosslyn has done a lot of experiments on mental imagery. And his work is controversial in some regards. But let me see if I can think of a good test case for this. I want you to imagine this cup. And I want you to close your eyes and imagine the rotation that it goes through as I turn it this way. . . . and then you can see the handle coming out on the other side from behind the cup. I want you to imagine all this in some detail.

V.V.: I’m doing it badly.

D.D.: Well, okay. But the fact is, it's hard. The fact is if you really have images, it shouldn't be hard.

V.V.: Why? I have images but they are weak, faint.

D.D.: All right, there are weak and faint images?

V.V.: Yes, of course.

D.D.: And they have a property that no regular image has. Namely, they can simply not go into some topic at all. Imagine a pirate of the Caribbean. Did he have a wooden leg? Well, now, if you imagine him, you don't have to imagine him with a wooden leg or without a wooden leg. You don't even have to get to the question of his legs in order to imagine him. But if you draw a picture of a pirate, the picture has to show him with a wooden leg or not with a wooden leg or it will have to hide his legs behind something or otherwise obscure the issue. A picture of a (whole) pirate can't just "fail to go into the matter" of whether or not he has a wooden leg. But a description of a pirate can be as incomplete as you like.

D.V.: Give us time for creation of detailed images, more time.

D.D.: Yes. I should give you a little more time for creation. And time to imagine the whole pirate standing on the shore with an old-fashioned pistol in one hand. Now that you've done that, imagine him in a lot more detail. And then I can start asking you questions like: can you see his belt? Does he have a belt on? How about his shoes or bare feet? And you will realize that you just haven't gone into that. But if you draw a picture that meets the minimum demands for being a picture of a pirate, you either have to obscure that area or you have to settle it. In fact this is one of the important things about images. I learnt this in a very interesting way when *Consciousness Explained* came out. The BBC did a television documentary on it called, I'm sorry to say, "Mind Movies." It was a given that I would say, in the documentary, that there is no Cartesian theatre. We had a struggle with that throughout the whole production. The young producer of that program did a wonderful thing for me. I had been holding forth and beating up on various bad images of consciousness. And she said, "Okay, I see that those images are wrong. So what does a good image of consciousness look like—a good image of the architecture of consciousness?" And I realized that that was a perfectly legitimate question that I hadn't properly answered in *Consciousness Explained*. Then I developed a series of images of consciousness that I used in talks. If you have to draw something, there are a lot of questions that you have to settle, even if temporarily and defeasibly (and without making any strong commitments). These are questions that you *don't* have to settle if you just have a mental image of something.I think mental images are structures that never really require you to settle those questions. That's why they are not really like genuine, pictorial images.

D.V.: You recently published a book coauthored with Mathew Hurley and Reginald Adams, *Inside Jokes*. It's a fun book.

D.D.: Oh! I got you a copy. You already had it! You've already read it!

D.V.: It presents a comprehensive theory of humor. And this theory suggests that our sense of humor is based on our ability to construct mental spaces and to evaluate consistency within the contents of the spaces. So can you explain a little bit about those mental spaces and just briefly discuss the theory?

D.D.: The idea of mental space is really taken from the work of Gilles Fauconnier, a cognitive scientist now working in San Diego who used to be in Paris. He is concerned primarily with linguistics, and he is concerned with the framing of … well, of spaces. For instance if I start telling you a story then there is the story space. And if in the story John starts to tell somebody else a story, then we've got another frame or space. Or I can create another space by going into the mind of a character. We create these spaces all the time. Right now each of us has his own set of spaces that he is working in. And there are

spaces, outside these spaces, that contain each of them. These containing spaces, or our conceptions of them, may differ substantially from person to person. So we'd get into real trouble if we didn't compartmentalize ~~in~~ the space that we are now in. Fauconnier's main interest in creating this theory was in figuring out {figuring out = getting to understand] how things like demonstratives work, for example the demonstrative "that guy." You can say "that guy" without pointing, and everybody in the space knows whom you're talking about even though you have not obviously specified a person very precisely. The spaces solve problems like that.

Well, what Mathew Hurley realized (his model is the heart of the book) is that one of the roles that a mental space plays is that of a lobby, like the security clearance area downstairs in a building lobby where people come in and have to be checked out, to make sure that they have the right to be there. You don't trust everybody who wanders into the building. So there has to be a sort of testing ground for stuff that comes in from your senses and from the communications which you have. The brain is engaged, all the time that you're awake, in furnishing the current mental space, correcting its furnishings, adjusting here and there and anticipating what else is there, sort of filling in the blanks. And it's under time pressure as it does this. It's important that you stay current with what's going on. So the brain gets sloppy, makes a lot of jumps to conclusions, sometimes makes mistakes. And if these aren't corrected, then you have in effect a security problem, you're going to contaminate your world knowledge, you're going to let in false information and cripple your brain's ability to make sense of the world. So the brain needs to have a security system, a system of sentries or guards that are constantly checking on these points. That's expensive. The brain can't get that for free. So the brain's solution is to create a reward system that rewards the guards in effect for playing this role (it's a little example of pleasure, an example of reward), for detecting the hidden false assumptions, the contradictions in the mental space.

When you detect a contradiction, either you immediately resolve it by creating another mental space where that is handled, or something else has to happen. If you detect a problem—I say "you," but really it is the brain—if your brain detects a problem that it can't resolve, then there is anxiety: "uh-oh! uh-oh! Something's wrong, something's wrong!" And this is phenomenologically familiar to us. We're walking in the kitchen and something's wrong ... with what? What is it? You just have the sense that something's wrong, but you don't know what yet. Interestingly, in English there is an idiom '"hm … something's funny here" or "what's a funny smell". And in this context it doesn't humorous, it means worrying, it means "uh-oh, uh-oh." Then if you get resolution, if you say "Ah!" and you see what the problem is, and if seeing that comes at the right time lapse, very quickly then that initial momentary "uh-oh" creates a sort of springboard. I call it the Huron trampoline because David Huron has a similar theory of music and the brain. [David Huron, *Sweet Anticipation: Music and the Psychology of Expectation*, MIT Press, 2008] The tiny micro-anxiety detecting an error then intensifies the little jolt of pleasure that comes right after it, when you realize that "Uh! Everything is all right."-So the basic moment of humor in the model which we defend is the discovery of a potential source of anxiety. This source of anxiety occurs whenever you make a mistake that's a cause for concern. Humor arises when the mistake is swiftly resolved and you then get the little "aha." What happens in humor is closely related to the aha phenomenon, the joy of sudden discovery, but we think it's also significantly different. But that's why when people solve a puzzle or a riddle they sometimes laugh; they find it sort of amusing ... sort of. So there is a sort of grey area between humor, as here understood, and problem solving, and it's probably no accident that the pattern of humor that we note goes along with jokes and riddles. Riddles are the cheapest, most easily formulated type of humor. Children catch on to riddles very early; they love them. But our theory applies also to much subtler forms of humor.

D.V.: What about the Phi-phenomenon? Why isn't it funny? When you discover there is a motion and you suddenly know that there is no motion, why wouldn't you laugh?

D.D.: First of all I think that with a little setting we could probably make a phi-phenomenon that is amusing. Think back to the earliest days of cinema when the projected images of an oncoming train made people dash out of the theater. In those days maybe the phi-phenomenon would be genuinely upsetting, especially on a large scale. And people might laugh when they realized that they were just fooled by this illusion, this illusory motion. The extreme effects in three-D movies often evoke laughter today, but we are too accustomed to ordinary cinema and video to make the temporary mistaken *commitment* that humor depends on.

One of the challenges that we put to our readers is this: we've provided a quite careful and explicit list of conditions for humor. If you meet all these conditions, you have humor. If you don't, you don't. So that now leaves us wide open to two kinds of counterexamples: phenomena that meet our conditions but manifestly aren't humorous, aren't funny; or of course the alternative, something that is hilarious but that doesn't meet one of the necessary conditions that we argue for. We haven't had to revise the model yet. We are waiting.

D.V.: You actually use the word *'qualia'* many times. But what do you mean by *qualia*? And why do you use that technical term now?

D.D.: I suppose I was just deciding that you don't want to fight all the battles at once. I hope I've made it clear that I think that in a sort of technical sense the notion of qualia is incoherent; but in a *loose* sense, used just to refer to features of subjective experiences, it's acceptable shorthand. Wilfrid[1] Sellars makes the distinction between the manifest image and the scientific image. The manifest image is everyday stuff. That includes colors and sounds, and free will, other minds, possibilities, dollars and rubles, tunes, and so forth. These are all kinds of things that are very hard to identify or characterize in terms of atoms, molecules and the other elements of the scientific image. Sellers says (and I simply agree with this point) that philosophy considered abstractly is the attempt to say how things, in the broadest possible sense, hang together, in the broadest possible sense. How do the manifest things - colors, smells, tables and chairs, tunes - how do we fit them together with the ontology of the scientific image? Nothing goes swiftly. People try. And you get eliminativism of one sort or another.

Right now for instance I'm really on the warpath about free will, because there's a veritable chorus of neuroscientists who say that free will is an illusion. I want to calibrate your view if you make this claim about free will. I want to know: according to you, is solidity an illusion, or are colors illusory or are dollars an illusion? Is poetry? Just tell me what your answer to those questions is, because I want to know what you are saying when you say that free will is an illusion. Simply declaring a phenomenon illusory is usually not a good way of resolving the difficulty between the manifest and the scientific image. Usually a more complicated but more nuanced story is better. It's the problem that I originally discussed years ago when I imagined that we come across these people who talk about "fatigues"; when we talk about the fact that we're tired, exhausted, they say they have fatigues, and want to know what "a fatigue" is. What should we do? It's not a very perspicuous ontology to declare either that "fatigues" in their sense clearly exist (whenever you're tired, you "have a fatigue"), *or* to declare that "fatigues" in their sense clearly do not exist! We do not want to be compelled to take fatigues in their sense seriously, taking on the burden of identifying "them" with states of a tired body. Nor, however, do

[1] Пишется именно так.

we just want to say '"Huh! There is no such thing as a fatigue," because in their world these "things" loom large. It is undeniable that they get tired the same way we do.

Okay, and so now in the above terms, if we use "qualia" just as a generic label for subjective properties without any more ideology, then of course there are many qualia. Pains are qualia, tastes are qualia, aromas are qualia, colors are qualia, and so on. But the term '"qualia" has been trumped up, beefed up and inflated by philosophers so as to turn it into a "term of art" or a "technical" term to which I think nothing at all answers. And so that's the sense which I think that qualia don't exist. But it's very hard to get people to take that issue seriously when it's so easy for them to shake their heads and say that Dennett hasn't noticed that sometimes human beings are in pain. Well, of course I have. It is just that I think that these people have a dead theory of pain, a notion of pain to which nothing really answers.

V.V.: Let's return to the topic of freedom. In your book *Freedom Evolves* and in some of your talks you claim that "If everything is determined, that doesn't mean that everything in our future life is inevitable, because as a matter of fact we avoid many things." Can you clarify why you believe that the fact that we behave so as successfully to avoid various things disproves the claim that every event that will happen in our future life is inevitable? Indeed, if some events are inevitable for us, then it's clear that some other possible events will be avoided by us—that's true by definition. But, if so, then the avoidance of some events could not be used as an argument for the non-inevitability of some events that will happen in our future. It seems that to show such non-inevitability, you must demonstrate that some events that will happen in our future could be avoided. And it is interesting to ask how it is possible to demonstrate this point.

D.D.: First of all, I think that you yourself see how difficult it is to articulate the thing that you are interested in here. Events that *really are* in our future are not going to be avoided because they will happen. That's what it means to say that they are in our future. People sometimes say, "Well, you can't change the past but of course you can change the future." No. What does that mean? You can't change the future. The future is what will happen. But if you understand the future as the anticipated future—as what seems likely to happen, given current knowledge—then of course that is changeable. That's the whole point. Thinking about time and avoidance is not easy. But the first step is realizing that it's not easy because a lot of people just glide right over the problems. What is it to avoid something? Well, let's take a very simple example. I throw a ball at your head and you duck. Was the ball going to hit you? It was if you didn't duck. Your saw a ball coming toward you, the light bounced off the ball into your eyes and so caused events in your brain which triggered off the ducking motion, and the ball sailed over your head. That's a paradigm case of avoiding something. Is it consistent with determinism? Sure. You were determined to avoid that ball.

V.V.: But it seems to me that it is also consistent with the idea of the inevitability of our future because if some events are inevitable, then so, by definition, some other events will be avoided.

D.D.: Why so?

V.V.: Is some events are to happen to us, then some other possible events will be avoided by us.

D.D.: Let's talk about things that are currently inevitable. Maybe I can imagine a definition of "inevitable" that will allow some events not to be inevitable indefinitely. In any case, here's an example of something that you might say is inevitable: I cannot be in San Francisco for lunch today. There is no way that is possible today. But suppose ... let's see, lunch in San Francisco ...and it's still early. You would not have to violate any laws of physics to get me to San Francisco by San Francisco noon. It is just not

possible *today*. Now that's a perfectly good sense of inevitability. There are other things which in that sense are not inevitable at all. No doubt there are a hundred restaurants in Moscow where we can have lunch, and none of them is such that it is inevitable, in the present sense, that we eat in that restaurant. The problem is *not* that there is one of these restaurants that is such that we will end up eating in it no matter what we choose. Rather, where we will eat in Moscow is determined by whatever choice of restaurant we come to make. However, in the San Francisco case, there is no path now that will lead to a situation in which Dmitry and I will have to chose whether or not to fly to San Francisco for lunch. But if we now choose to eat at a certain restaurant in Moscow, we are not *avoiding* eating in San Francisco, because present circumstances do not determine that the situation arises in which we have to choose whether or not to go to San Francisco for lunch. So the inevitability that we won't eat in San Francisco doesn't imply that our not eating there is something that we will avoid through any choice that we ourselves make. This point has nothing in particular to do with determinism.

And here is a further way of thinking about this. Here is a truism: *if determinism is true then every event in my future is determined.* That's true by definition. So do you now want to infer, from this truism, the second claim that *if determinism is true then every event in my future is inevitable*? This second claim doesn't follow from the truism. I mean, if it does follow, then tell me what does the second one add to the truism? The truism is trivially true. How do you get from the first one to the second? There is no path. And one way to see this point is to notice that the word "inevitable" (meaning what will happen no matter what I choose) marks one half of a distinction for the other half of which English supplies no good word. The word "avertible" (meaning something that *can* be turned aside through our choice or action) exists in the dictionary, but nobody ever uses it.

V.V.: But it did exist as far as I know.

D.D.: Yes; well, "avertible" is just "avoidable." But there is another contrast. We want to avoid bad things. We want to choose good things. There is no good word for the opposite of "avoid." The word for the opposite of "avoid" is something like "seek" or "get." Now let's just say it's *get*. We want to avoid the bad and get the good. Okay, now consider the following sentence: *if determinism is true, then every event in my future is either gettable or is ungettable*. What does that sentence mean? Not that I'm going to get some things and I'm not going to get other things. That's true whether the determinism is true or false. But, for the same reason, it doesn't follow that if determinism is true I will never be able to get anything. Similarly, it's false that If determinism is true, then I will never be able to avoid anything. But if I am able to avoid things that will be in my future if I do not make choices that prevent them from occurring, then my future is not inevitable in the relevant sense, is it?

D.V.: In your books you sketched briefly your theory or concept of morality. For instance in *Darwin's Dangerous Idea* you have a chapter devoted to virtues of morality, and in *Breaking the Spell* you have a chapter devoted to relationship between religion and morality. Can you express briefly what would be your main ideas about ethics, morality, and what is good?

D.D.: Yes. First of all let's start with what *isn't* involved and that is: we don't need God or religion for morality. Maybe there was a time when we did, but we don't need it now and in fact for several millennia people have figured out what is right and wrong and then very conveniently said that that's what God told them to do. Nobody would live with Old Testament morality today. It would be viewed as a nightmare. We've learnt that slavery is bad, that wife beating is bad, that many things that not only are tolerated but also are encouraged, commanded in Old Testament morality are now viewed as crimes. So morality has evolved over time, and how has it evolved? It's evolved because people have engaged in that most human and wonderful activity of mutual persuasion—they got together, they

talked it over. And they've convinced each other that certain values are pretty much shared. If they weren't shared initially, they've come to be shared.

We've expanded the circle of agreement, and it's based on some very fundamental virtues that Hume described as *natural virtues*. Then we have the *artificial virtues,* which are more specialized and have been added on as with we've civilized ourselves. Thus we've learnt the importance of truth telling, of integrity and gentleness and tolerance and so forth. The ideal is not yet achieved, but we can imagine an ideal of a communal activity of mutual persuasion that allows a group to settle that this is how we, the members of the group, are going to behave. We arrive at principles that *just about everybody endorses*—what could have a greater warrant for us, what could have a greater command on our allegiance, than such policies about how we should behave? This is our sense of our morality, it's human based, it doesn't get imposed on us by God or by anybody else. For the same reason, we wouldn't let Martian space pirates tell us how to behave. We shouldn't let any imagined religious figures tell us how to behave. We'll figure out how to behave, we trust our intellectual equipment and our inbuilt sense of affection and friendliness and a sort of default pacifism which I think runs pretty deep. And that's the best we can do for morality, and it's plenty good enough. I think Steve Pinker's *Angels of our Better Nature* is an excellent book, very well done. Pinker is a very smart fellow, and I like his view of morality a lot. I have one quarrel of emphasis, and that is this. It is true that violence is now way down, way down; and we are getting better and better about not doing many of the most horrific things that people do with each other. Pinker is right, here. And it's stunning because it's so initially counterintuitive, but he has a good account both of why it's true and of why it doesn't seem to be true. But I think he doesn't devote enough attention to questions about subtle oppression and manipulation. Right now we are not marched at spear point to work. But many people are in a different kind of slavery just to stay alive, and they are exploited by bosses or by the system. I think we should recognize the subtle thwarting of our desires and of what would make our lives better. I think many lives are blighted, are cramped, are pinched, by features of society whose harmful consequences we very much underestimate.

D.V.: To turn to the value of life, a few years ago you published an article (and you also had a presentation) about what you called a Full Body Apoptosis. I think it was a surprising reply to the request for new ideas about how we can enhance human beings. So can you explain in this context your argument about what do you think makes human life valuable?

D.D.: *Artifact* is a design magazine that wanted to do a special feature on imagined adjustments to human beings, *Homo sapiens* 2.0. Contributors were invited to defended proposals; it had to be something that was technologically feasible in the near future and something that you would submit to yourself. So do I want an eye in the back of my head, do I want eyes in any of my fingers? What other changes would I ask for? A huge memory? Who knows? Well, I decided I would concentrate on death. And especially today in the developed world most of us can anticipate a future in which we might well spend months or years gradually turning into vegetables with tubes up our noses, lying in bed without a mind, deteriorating. iIt's difficult to prolong life, as is well known; it's expensive, but few people seem to be achieving any *meaningful* extension of how long they live. Suppose that you ask many people how they would like to die. Would you rather be struck by lightning, or would you rather have this long lingering process? I think that they would say that they would prefer to be struck by lightning. But it's very hard to arrange to be struck by lightning. And, besides, you don't want to arrange it. Rather, here is what you want. You want to die as swiftly and suddenly as possible, consistent with the safety of those around you. And you don't want this to be by default, a matter of somebody (you or your family, or the state) judging that your life isn't worth living any more. This is the important point. So I envisaged—let's

just call it *a pill,* but there are various ways it could be devised—something that you take that is slow acting and impenetrable, and at some point between 80 and 85 or ... ?

D.V.: Between 85 and 90…

D.D.: Sometime between 85 and 90 you just simply drop dead— 'Boom!'—just the way you wanted to be hit by lightning. You can't predict it. Immediately it was suggested to me that you would want to build in a twenty-minute warning. So you can get off the highway, you can pull up your pants, you can call your children, whatever. So let's build that in, it's all you have. Very short. I say, "Half an hour maximum. No more." So at some point between your eighty-fifth and ninetieth birthday suddenly the alarm will sound and you will know that in half an hour you are going to fall over dead. Make whatever preparations you want for that. If you like this idea, the price you have to pay for it is that it may cut you down when you are still healthy.

V.V.: That's a problem?

D.D.: The only alternative is a system which many people think they prefer, but which I try to show them that they really shouldn't prefer. That's the system where they get to prolong life indefinitely until they decide that it's no longer worth living. But what that means is that they will end up spending the last part of their lives measuring the value of their own life constantly. "Mmm. Am I feeling good enough today? I want to know … I think I might enjoy watching one more television game show." And of course everybody around is going to be looking at their watches and wondering whether grandpa's life still worth living. An old friend of mine, philosopher Kurt Baier, once said "Socrates says that the unexamined life isn't worth living. The overexamined life is nothing to write home about, either." I think it's a wonderful comment on the overexamined life. I want to avoid the overexamined life. It's very interesting in today's world, especially with Twitter and Facebook. A lot of people don't seem to realize this, and they seem to think that the point of life is to have your life under constant review and analysis and evaluation, as if you can't act without asking yourself "Where on the scale of all the thrills that I have had in my life should this event rank? Is it the best, the second best, the seventieth best? Is this person a ten or a nine and a half? Does doing this rank eighth on the 'hurrah!' scale?"

If you go through life evaluating everything, if that's all you are doing, then that's all you are doing. And that's a terrible way to spend your life. Especially at the end of your life, you shouldn't be oppressed by the issue of whether your life is still worth living. You should be just living your life and let the lightning strike when it does. At least that is my conviction, so I describe the system which I would submit to. But I pointed out in the article that I would not want to impose this on anybody. So if a consensus could not be brought about on this, then we should just stop with the system we have. I would not want to impose my system. It's very tricky. Maybe we are stuck in a consensual bind {<= "bind" is colloquial American English for "predicament"} that we can't escape from. And then individual people are just going to have to take the problem upon themselves. But I think in fact that it's going to be a serious problem. I think that people are already beginning to feel the pinch, the pressure. It affects all the people who are confronting the economic fact that they have an estate that they would like to leave to their children and grandchildren, an estate which will be completely exhausted through its use to pay for them in their eventual vegetable state unless they do something about this problem before then. I think something like seventy five or eighty percent of all the health-care costs in America are spent on the last six months of life for people. I mean that's obscene, that's ridiculous. And so there is already social pressure, it's not a pleasant social pressure. I think all people are beginning to feel that once they're old, they're lived out, they're not welcome. That's not a nice way to feel. But we are going to have to realize that if we don't want to feel that way, then we could arrange something which takes our

deaths out of our hands and out of our family's control, out of the state's control. You can't arrange to be struck by lightning.

D.V.: You are a well-known atheist. You offer a lot of arguments fighting religion. What do you think about life after death? Is it possible, according to your theory? Can the mind survive the death after body?

D.D.: First of all I think that it's almost embarrassingly clear that life after death is wishful thinking. It is the most natural and seductive idea, it's no wonder that everybody thinks of it and everybody loves the idea, especially if it falls to you to have to explain to a child about the death of a parent or of a loved one. The capacity, the opportunity to say "she is still in heaven, she is looking down on you"—this is tremendously consoling. I think there is no mystery about why it is such an appealing idea. Now, is it possible? I've said in one sense "Yes, it's possible, it will be soon possible," because what you are is the information in the dance which your molecules are dancing. We can replace every one of those molecules and you will go right on living. And if we could just store the dance which is information then it could be put in your new body in the same way you can have a new edition of *Ulysses*, make a new copy of a song. So we could make a new copy of you. It should be possible in principle.

V.V.: But what about the personal identity in such a case?

D.D.: In that case I think the identity goes with the dance that goes with the information. Now we have these puzzle cases about which of several candidates for you will really be *you*. And there is the so-called *closest continuer theory* which says that the one that will be you is the one that is most like you. For my part, I don't have any trouble with the idea that personal identity is not importantly different from the identity of the ship of Theseus. I mean the famous example in which Theseus[2] goes off in his ship. Then every plank, every board, everything is gradually replaced, and it's still the ship of Theseus. But what if somebody has been going along behind the ship picking up all of the discarded pieces, and that person then reassembles those pieces somewhere, patching them up a little bit? Isn't that reassembled ship also the ship of Theseus? Well, at this point you may think that this is a question for the insurance company and the law. We know all the facts about the ship that has been gradually rebuilt and also all the facts about the ship that has been built out of the discarded parts. And we discover that the facts do not clearly tell us which way to go. The facts do not immediately make it clear that *this* ship rather than *that* ship is *the* ship of Theseus. Okay, but in the case of personal identity, we also have the sense that it's not like this mysterious problem of ship identity. However, I think we have to come to the same realization about personal identity that we did about the ship. The hopeless idea that there is a little magical thing which is either there in one of them or is there in the other, and this thing determines which is the real ship or the real you. But we just have to abandon that idea. There is no reason to believe in that sort of entity. It's very tempting to believe in, but there is no reason to.

V.V.: One or maybe two personal questions. In one of your recent talks you mentioned a philosophy of language class of 1960 with Quine as professor and you, David Lewis, and Saul Kripke as participants. Can you tell us some more about your relationship with these famous people?

D.D.: I was an undergraduate. The others were a little bit ahead of me, more than one year ahead of me, I guess. But the graduate students were Tom Nagel and Davis Lewis, Gil (Gilbert) Harman and a number of others—those perhaps are the best known. It was a very exciting thing when I arrived at Harvard in 1960. I was already very interested in Quine from having read his book *From a Logical Point of View*. And

[2] http://en.wikipedia.org/wiki/Ship_of_Theseus

he was publishing a brand new book called *Word and Object*, and this was the first time he taught from that book in his philosophy of language class. So I was very glad to be there and I was very impressed with the fast company. It was a very strong bunch of students. I got to know some of them pretty well but others not very well because I was an undergraduate and they were graduate students. But I got to know David and Stephanie Lewis pretty well and Gil Harman later. I never got to know Saul Kripke well. We hardly conversed. Our paths just don't cross that much.

V.V.: You told the story about the gold Rolls Royce of Montague and David Lewis.

D.D.: Oh you heard that! Yes, it's right. Richard Montague was the logician at UCLA, a brilliant and eccentric man—and a rich man. He lived a very flamboyant life. He was eventually murdered by a gay hustler he picked up. Nobody ever was convicted for that crime. He lived in a beautiful house up in the hills outside Los Angeles. And he had a gold Rolls Royce. I didn't know that when he and David Lewis came down to our colloquium at the University of California at Irvine,[3] in the early years of Irvine. The place was brand new. There was hardly any grass, little spindly trees, and we walked out to the parking lot in the evening after the colloquium. And there in the parking lot I saw - in Irvine(!) - this gold Rolls Royce, and I had never seen one before (and have never seen one since). I just couldn't keep from exclaiming, '"Oh my God! Look at that! Look at that! It's a gold Rolls Royce!" And "What on Earth can that be doing here!" And then I had to go and, you know, stick my nose up against the window, walk all around it, and all the time of course exclaiming about the tastelessness, the ostentation of owning a gold Rolls Royce. And David just kept his lip buttoned and so did Richard. Then after I had had my full of marveling at this car, Richard got out the keys, got in, and they went off. I felt very foolish.

D.V.: Maybe the last question. What you consider now as the most promising areas for philosophical investigations?

V.V.: And what are you currently working on?

D.D.: I will do the second one first. I'm about to publish a book on *Intuition Pumps and Other Tools for Thinking* which goes through my work and pulls out the best parts and then changes them all. I mean we are working on each one to make it more accessible, more portable, you might say. And also criticizing intuition pumps that I think don't work well. I have had a lot of fun doing that. And it should be in the hands of the publishing house soon. I think it won't be out till after Christmas, but we'll see how it goes this summer and fall. But then I have several projects that I want to turn to. One of them has to do, as so often before, with what you might call the *computational architecture of consciousness;* I've taken on some new ideas which I haven't fully digested, I haven't figured out quite out to handle them. One of them is the Bayesian predictive-coding point. Another is making more sense of the idea that although the brain is a computer of sorts, the parts that it is made of are very, very unlike the parts that my laptop is made of. In the brain, there are two hundred billion extremely individualistic neurons, and there are no two alike. And in fact I've become impressed with the fact that every neuron in your brain is a direct descended of single-cell organisms, single-cell eukaryotic organisms sort of like amoebas. These organisms fended for themselves for a billion years on this planet, so they have a lot of knowhow, a lot of competence in their ancestry. I think that it can be useful to start thinking of the neurons in the brain as a kind of imprisoned work force, composed of agents with agendas of their own, and they have to work to survive. Sometimes I joke about this: I say that the traditional model of computation is a sort of Marxist world of *from each according to his abilities, to each according to his needs* and you never have to worry about needing more things, you always get a meal, you always get to use the energy you

[3] Город в штате Калифорния, где Деннет жил с 1965-1971гг.

need, whereas the brain is a more anarchic situation where each individual neuron *does* in effect have to worry about staying alive and increasing its influence. And when you start thinking of neurons as little selfish agents that are living in these very restricted circumstances, you've got a very different idea of computation. So I'm trying to work out this idea. I have some colleagues, post-docs, who are helping me think about this. We jokingly call this *brain wars* because we think that the turmoil in the brain really is a kind of serious competition, it goes on and coalitions form and détente is possible, but there are also unresolved conflicts and so forth. So I have to work out some more implications of that. It might turn out to be a crazy idea but for the time being I want to push and see what I can make with it.

V.V.: Thank you!

D.V.: Thank you very much!

D.D.: I enjoyed it.

Transcribed by A. Kuznetzov, edited by R. Howell, and authorized by D. C. Dennett