BOLD CONJECTURES

Conjecture Institute

FIRST EDITION

Digital ISBN: 979-8-218-78566-6
Print ISBN: 979-8-218-78567-3

Interviews by Naval Ravikant, Tyler Cowen,
Sean Carroll, Robert Lawrence Kuhn, and more

VOLUME I

BOLD CONJECTURES

Select Interviews of DAVID DEUTSCH

EDITED BY
LOGAN CHIPKIN

Dedicated to my parents, Joel and Robin Chipkin.

Thank you for your love. And for your patience.

Major support for this volume
was provided by Naval Ravikant.

Additional support by Darren Wiebe, Amaro
Koberle, Tomas Geerkens, Ayush Modi,
Anders Hvelplund, and Dirk Meulenbelt.

Conjecture Institute gratefully acknowledges
their generosity in advancing this work.

CONTENTS

PREFACE

Physicist and philosopher David Deutsch's two books, *The Fabric of Reality* and *The Beginning of Infinity*, offer a deep and coherent worldview that has improved on humanity's ideas in physics, epistemology, morality, aesthetics, and other fundamental domains of knowledge. He's been interviewed over one-hundred times during his career, giving his readers hours of additional content that elucidates and expands upon the ideas in his books, in addition to ideas far afield from his writings.

David's interviews provide more than enough content for a book—and that is precisely what we've done. *Bold Conjectures, Volume I: Select Interviews of David Deutsch* is a compilation of over a dozen interviews of David Deutsch.

We have transcribed about seventy interviews that did not make it into this publication. Fortunately, the *Bold Conjectures* series will not end after its first volume.

To be sure, the *Bold Conjectures* series will not only consist of David Deutsch interview transcripts. For example, a subsequent volume will consist of original essays by select physicists about various topics, written for a lay audience. Ideally, many future Bold Conjecture volumes will follow the spirit of John Brockman and consist of original essays by world-class thinkers.

As you read this compendium, please remember that you are reading transcriptions of extemporaneous speech. Therefore, you may encounter imprecisions, accidental contradictions, and awkward sentences that sound acceptable in audio form but less so in written form. This does not reflect on any of the speakers whose words I have transcribed.

Please note that David Deutsch's views may have changed since the times of the transcribed interviews found in this compendium (for instance, his views on free will, time travel, the nature of constructor theory, and other things have evolved).

Brackets indicate a word or phrase that I couldn't make out from the interview audio, or else that I slightly modified for grammatical purposes.

Interview dates should be considered approximate.

Any errors you find in the transcript are due to me.

Yours,

Logan Chipkin

President & Cofounder, Conjecture Institute

ROBERT LAWRENCE KUHN:

CLOSER TO TRUTH

About the interviewer: Robert Lawrence Kuhn is the creator, executive producer, writer, and host of Closer To Truth, the PBS/public television series on Cosmos, Life, Consciousness, and Meaning that presents leading scientists, philosophers, and creative thinkers discussing fundamental questions. Dr. Kuhn has published over 30 books and is a renowned China expert, international corporate strategist, investment banker, and public intellectual.

About Closer To Truth: Closer To Truth is a broadcast and digital media not-for-profit organization and series. On the air continuously since 2000, the weekly, half-hour television show airs on over 200 PBS and public TV stations. Closer To Truth is created, executive produced, hosted, and written by Robert Lawrence Kuhn, and co-created, produced, and directed by award-winning filmmaker Peter Getzels.

Closer to Truth homepage: https://closertotruth.com/

Closer to Truth: David Deutsch homepage:
https://closertotruth.com/contributor/david-deutsch/

The following nine episodes were recorded during a single interview session with David Deutsch. The first segment was released in Season 10 (2012), with additional segments appearing in later seasons.

WHAT DOES QUANTUM THEORY MEAN?

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - What Does Quantum Theory Mean?
Description	Quantum theory may be weird—superposition and entanglement of particles that in our normal world would make no sense—but quantum theory is truly how the microworld works. What does all this weirdness mean? How to go from microworld weirdness to macroworld normalcy? Will we ever make sense out of quantum mechanics?
Link	https://youtu.be/0CwPa0tScf8
Ideas	• In the approximate theories that we think might be like quantum gravity, there is no fundamental difference between different times of the same universe and different universes at the same time. • There seems to be no limit to how finely the multiverse can subdivide itself. • The apparent flow of time is an emergent property of the multiverse.
Topics	differentiation across the multiverse • multiverse • number of distinct universes • quantum gravity • quantum interference • time

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Robert Lawrence Kuhn** David, it was difficult enough for me to hear about the multi-world interpretation of quantum theory with its continuous branching of innumerable different universes. Now I'm told that all these different universes have always existed and always will exist, and it's very difficult to understand.

0:21 **David Deutsch** What's happening to them, and the reason that they used to be thought of as branching or splitting, is that they become differentiated from each other. At the beginning of time, perhaps the Big Bang or whatever was the beginning of time, the universes may all have been identical, and physical processes happening in them caused them to differentiate according to the laws of motion of quantum mechanics. So the total number, as it were, remains constant, but the degree of differentiation between them increases very rapidly, and that's what we call the arrow of time.

0:58 **Robert Lawrence Kuhn** Well that is unbelievable. Now, what kind of number can express the total number of them? Is it a finite number?

1:04 **David Deutsch** Yes. We can't answer that question definitively because it depends on quantum gravity, which is a theory that we don't have yet. But there is very good reason to believe that the total number of different universes is finite while the actual total number of universes is infinite. So they keep dividing up among themselves and there's no limit to how finely the multiverse can subdivide itself, but the total number of distinct different universes with different contents is a finite, though enormous, number.

1:41 **Robert Lawrence Kuhn** And as they differentiate from each other, what happens to all of them? I think I'm living in one.

1:51 **David Deutsch** On the gross scale, most of the elements of human experience, once they have separated in what we usually perceive as a random event like a coin toss or winning the lottery, they no longer interact, or rather their effect on each other becomes exponentially small so that it's immeasurable, but it never goes away. It is in fact always there to some very, very tiny degree, but the smaller the scale you look on, the more important these interactions between different universes become.They are what is called in quantum theory 'interference processes,' because the universes interfere with each other.

2:37 **Robert Lawrence Kuhn** Help me to understand the difference between the universe and the multiverse, in that the multiverse can have many universes in it, but it has more than what's in the universe.

2:51 **David Deutsch** Yes. The multiverse is the name that this theory gives to the whole of physical reality. And it is a very complex object that is not described by the normal kind of numbers that we are familiar with, like three cats, three dogs, that kind of thing. But instead, it's described by mathematical entities that describe vast numbers of these objects at the same time, in the same formalism, and their interactions with each other.

3:23 **Robert Lawrence Kuhn** And when we select among them as they differentiate, is that a retroactive process? Because they were all there to begin with?

3:33 **David Deutsch** No, the tendency is that universes that have been identical become different.

3:40 **Robert Lawrence Kuhn** Because of the processes in each one?

3:42 **David Deutsch** Because of the processes, yes, in each one. Although really when they're identical we must regard that as a process across all of them. What's happening all the time as well, though, is that they rejoin. But they rejoin on a microscopic scale which adds up to, in our experience, a different set of laws of physics that we call classical physics and which can be approximated by these numbers that take only one value at a time. So on a microscopic scale, there are these quantum mechanical-type numbers that take multiple values simultaneously, and then they are approximated on the large scale by an ensemble, a lot of universes that look like classical physics and barely interact with each other.

4:35 **Robert Lawrence Kuhn** Okay, let's assume this is reality. What are some of the implications of it for the things that we think are very fundamental? For example, time. If there is a multiverse, what is the implication for the fundamental nature of time?

4:51 **David Deutsch** One of the exciting prospects about multiverse theory is that it sheds light on some rather notorious difficult problems at the foundations of physics, one of which is time, as you've just mentioned. So people often ask about the multiverse, "Okay, so there are lots of copies of me, which one is the real me? Which one am I? And how can there be more than one 'I,' since I have a unitary consciousness?" What people often don't notice is that this same question has been asked since time immemorial by philosophers wondering about time. Is the 'I' of ten years ago, who perhaps did something very embarrassing that I wouldn't do today, is that really me or is it not me? Or if a criminal repents, is it the same person and should that person be punished? These philosophical issues are all about whether the entities at different times are the same entity or not.

It turns out, amazingly enough, that, in quantum gravity, we don't have a theory of quantum gravity yet, but in the approximate theories that we think might be like quantum gravity, the different times, the snapshot of the entire universe all at different times, appear in the theory in exactly the same way that the different universes at any one time do. In fact, there's no fundamental difference between them. If we use relativistic transformations, we can transform the one into the other. And so there's no fundamental difference between different times of the same universe and different universes at the same time.

6:31 **Robert Lawrence Kuhn** That sounds remarkable. That sounds like a radical transformation of the classical Einsteinian four-dimensional block universe, which sort of sits all in the same four-dimensional space.

6:43 **David Deutsch** It is radically different, although it shares some things in common, but we don't know how to integrate them yet. But it seems to me that this is obviously a clue to how to integrate Einstein's relativity with quantum theory, which is an unsolved problem.

6:58 **Robert Lawrence Kuhn** So to state the remarkable thing again, which you say in normal language, but it's so startling that different universes in the multiverse will deal with different times of the same sorts of events.

7:12 **David Deutsch** It's not that they deal with them, it's that the different times, that is, the universe yesterday is the same kind of object as an alternative universe as we might call it today. And in the way that this is described in these putative theories of quantum gravity, there is no fundamental difference between those. It's like one of them is different universes in the east-west direction and the other one is different universes in the north-south direction.

7:44 **Robert Lawrence Kuhn** And these are existing with some sort of, I don't want to say co-temporally because that may confuse everything, but these are all existing, let me just say.

7:54 **David Deutsch** They're all existing on the same basis as each other. None of them is privileged relative to the other. As you rightly say, the flow of time is an emergent property of this thing, and we can't think of the multiverse as being in time, it's more that time is a feature of the multiverse.

WHAT IS ULTIMATE REALITY?

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - What is Ultimate Reality?
Description	What is the deepest nature of things? Our world is complex, filled with so much stuff. But down below, what's most fundamental, what is ultimate reality? Is there anything nonphysical? Anything spiritual? Or only the physical world? Many feel certain of their belief, on each side of controversial question.
Link	https://www.youtube.com/watch?v=GBc6vj5-wko
Notable	• The four fundamental theories that comprise the fabric of reality are: quantum physics, the theory of evolution, the theory of computation, and the theory of knowledge. • The theory of evolution is the basic theory of emergent properties. Darwin solved one of the fundamental mysteries of nature, but his theory cannot be expressed in terms of atoms. • Of all of the possible transformations that are permitted by the laws of physics, the overwhelming majority only happen if the right knowledge is present. Knowledge is therefore a fundamental entity.

Topics Darwin • emergent properties • fundamentality of an idea • genes • information • quantum physics • the four strands of the fabric of reality • the theory of computation • the theory of evolution • the theory of what knowledge can do • theory of relativity

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Robert Lawrence Kuhn** David, I've always wanted to understand ultimate reality. Sometimes that's a silly idea to be able to think that you can do something like that. Your book, *The Fabric of Reality*, really made a profound effect on me because it showed that there is progress that one possibly could make. So how did you do that? What are the elements of the fabric of reality?

0:24 **David Deutsch** It's about what's fundamental, and more specifically, about the most fundamental things that we know. I don't attempt to find ultimate reality because I think that's a kind of chimera. We always should start with problems and start with what we know and then try to make it more fundamental than that. And what I call fundamental, a fundamental idea, is one that is needed in the explanation of many other ideas or many other phenomena and so on. And the most fundamental ones we know are basically the ones that are needed in the explanation of practically everything. And I wrote *The Fabric of Reality* because I realized that the most fundamental theories—there are four that I picked out as being the most fundamental—formed a sort of unified fabric of reality, a conception of the world, where none of them could be understood without the other three. And they were: quantum physics, which is my actual field, and then the theory of evolution, the theory of computation, and the theory of knowledge, which is usually not even considered part of science, but those were the four strands.

1:42 **Robert Lawrence Kuhn** Now just seeing that for the first time, one's impression [is that] these are each important things but they're radically different categories even in type or orders of magnitude, and so it sounds like they're all interesting but they don't necessarily fit together.

2:00 **David Deutsch** Yes, they seem at first sight to be different kinds of [things]. Three out of the four seem to be tied to us humans or to life on Earth or something, while the fourth is universal. But the closer you look at these four theories, branches of knowledge, strands of the fabric of reality, the more you realize that you can't understand any of them without the others. They're all intimately related.

2:30 **Robert Lawrence Kuhn** Give me a quick synopsis of each of the four.

2:33 **David Deutsch** Quantum physics is one of the two fundamental theories of physics. It is the language in which all other theories are written. The other fundamental theory is Einstein's theory of relativity, which describes space and time. Quantum theory is the language that all other theories in physics are expressed in, and it sort of constrains the kinds of ideas that one can express within physics. It's the deepest and most successful theory.

3:08 **Robert Lawrence Kuhn** Evolution?

3:08 **David Deutsch** The theory of evolution is the basic theory of emergent properties. It's how large objects can be understood in terms that do not follow from their low-level definitions in terms of atoms. And so we have laws like the principle of evolution which is a rigorous law of nature and in terms of which Darwin solved one of the fundamental mysteries of nature but it cannot be expressed in terms of atoms. So that's the theory of evolution.

3:52 **Robert Lawrence Kuhn** Computation?

3:53 **David Deutsch** And then theory of computation is the theory of what processes in nature are independent of or transcend the material substance that they are embodied in. So for example, I can say I had an idea last year and now I'm telling it to you. And that idea is an abstract entity that is first of all instantiated in the brain, then it's instantiated in movements of my mouth, then in vibrations of air molecules and so on. And it can be instantiated in ink on paper and an enormous variety of things. But in order to understand any of those transitions, you have to understand that what is affecting things, what is moving things here, is the information itself, not its instantiations. And the general theory of how information is processed in the world is the theory of computation.

4:57 **Robert Lawrence Kuhn** And this does sound like it leads to knowledge.

5:01 **David Deutsch** And that immediately leads to knowledge, which is the kind of information that can do things, or solve problems as we would say at the human level. But in these terms, adaptations in living things are also a form of knowledge. So DNA embodies knowledge, human brains embody knowledge. Books and computers and the internet all embody knowledge. And the thing about knowledge that makes it fundamental is that if you think of any kind of transformation of a physical system, you know, from hot to cold or from a block of marble into a statue and so on, you think about all possible transformations that are permitted by the laws of physics, the overwhelming majority of those only happen if the right knowledge is present. So from the point of view of what can be transformed into what, it's practically all the theory of what knowledge can do, and that is why knowledge is a fundamental thing in the physical world.

6:08 **Robert Lawrence Kuhn** Does knowledge have to have a purpose, a teleology, a meaning in order for it to be knowledge?

6:17 **David Deutsch** With the benefit of hindsight, one can determine that it has a meaning because it's the meaning that keeps it in existence. One way of expressing the fact that knowledge is information that does something is that it's the kind of information which, once it is embodied in a certain type of physical system, it tends to remain so. And in biology, that happens because its rivals die.

6:45 **Robert Lawrence Kuhn** DNA is a good example, representing the fitter of the species.

6:50 **David Deutsch** Exactly, or rather the fitter of the genes. And with human knowledge an idea remains embodied in things like books and brains to the extent that it does something, like it enlightens people or it allows the physical world to be manipulated or whatever, but it has to have some kind of use.

7:12 **Robert Lawrence Kuhn** So integrating those four, what are the implications? Where can you take it forward?

7:20 **David Deutsch** In my first book, in *The Fabric of Reality*, I just wanted to say that these four strands are fundamental in this sense, and I wanted to say that you can't understand any of them without the other three, and that all of them have been kind of underestimated in that they have been accepted as the right explanation in their own field, but they haven't been taken seriously as a component of people's worldview. So that, for instance, people accept quantum theory, but they don't accept its parallel universes implications. Other people accept the universality of computation, but they don't accept that this implies that it's possible to program artificial intelligence and so on.

8:11 **Robert Lawrence Kuhn** By integrating all of these into basically a theory of at least current reality, how do you feel about that now with some distance to your original creation?

8:28 **David Deutsch** It has made me think of the world in a much more unified way. And that's why I eventually came around to writing the second book, which is applying this fabric of reality to various issues which, on the face of it, like the four strands themselves, don't look as though they have anything to do with each other or with the four strands, and yet, the closer you look, the more integrated they are and the more they do have to do with each other.

MANY WORLDS OF QUANTUM THEORY

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - Many Worlds of Quantum Theory
Description	Quantum theory is very strange. No act is wholly sure. Everything works by probabilities, described by a wave function. But what is a wavefunction? One theory is that every possibility is in fact a real world of sorts. This is the Many-Worlds interpretation of Hugh Everett and what it claims boggles the brain. You can't imagine how many worlds there would be.
Link	https://youtu.be/Kj2lxDf9R3Y

Ideas

- Scientists invoke the cosmological multiverse theory to explain why unusual configurations of matter exist in our universe, while scientists invoke the quantum multiverse theory to explain outcomes of laboratory experiments (such as the two-slit experiment) and a number of natural phenomena (such as the solidity of matter).
- Some phenomena, like human thought and the outcomes of scientific experiments, are expressible in terms of universes, but other phenomena, such as quantum computations, are expressible only in terms of the multiverse.
- The quantum multiverse theory contains a lot more structure than just many parallel universes.

Topics

cosmological multiverse theory • many-worlds interpretation of quantum theory • number of universes • phenomena conditioned by multi-universe interactions • size of multiverse • two-slit experiment

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Robert Lawrence Kuhn** David, the many-worlds interpretation of quantum theory arguably is one of the most remarkable and astonishing claims anywhere in science. You've been one of the pioneers of this, so please tell me what it is and how can you believe such an extravagant claim?

0:20 **David Deutsch** What it is first: it is the idea that the physical world that we see around us, the room, the stars, galaxies and so on, is just one tiny sliver of the whole of reality. And the whole of reality includes many such objects, many of the kinds of objects that we have traditionally thought of as the universe. And so it's sometimes called the many-universes theory. I prefer to call it the multiverse theory because it contains a lot more than just those things that we used to call universes. It contains other structure as well. So the whole thing as a whole, reality as a whole, is the multiverse.

1:04 **Robert Lawrence Kuhn** Now how can you believe such a thing?

1:07 **David Deutsch** The reason we have to believe this, if we believe anything due to science, is…well, really there are two paths that force us to believe that there is a multiverse. One of them is simply to ask of quantum theory, which is the deepest theory that we have as physicists, in terms of which other theories in physics are expressed, we ask, "What does this theory say about reality?" And it turns out that in the equations of quantum mechanics, to express what happens in a process in the laboratory, you have to write out many paths for the apparatus, many different histories of it. And if this is applied to different histories of the laboratory as a whole, of the experimenters, of the world and so on, then that is the parallel universes interpretation. Historically, that is why people first believed in parallel universes.

I actually prefer a more concrete argument, which is that if you start with the experiments, if you start with a simple thing like the two-slit experiment where you pass a single photon through two different slits, that's already giving you a hint of the parallel universes kind of idea, that if you take seriously what happens in that experiment, the outcome cannot be explained by the idea that the photon passed through only one of the states or took any one path. Any one path would give a wrong answer. And the different paths affect each other. Again, if different paths affect each other, that means that different histories affect each other and so you can build the argument up into many universes.

3:08 **Robert Lawrence Kuhn** So first of all, I want to get a clarification. You used the term 'multiverse.' Most scientists today, particularly cosmologists, astronomers, use the term 'multiverse' in terms of generating multiple universes of the kind we know today through processes that involve part of cosmology, like inflation theory that shows how the Big Bang occurred and how the universe expanded, and branching off of other universes, and there are different other kinds. The multiverse you're talking about is radically different.

3:40 **David Deutsch** Yes, the cosmological multiverse theory is about universes that do not interact with each other. The only reason that cosmologists believe they exist is that they want to explain why unusual configurations of matter exist in our universe. They want to say that in most universes they don't. So that's the cosmological kind of multiverse universes that don't interact with each other.

4:12 **Robert Lawrence Kuhn** And that's not part of your thinking. You're agnostic on that, perhaps?

4:16 **David Deutsch** Yes, I don't know whether that's true or not. What I do know is that there's a lot less evidence for the existence of those than for the quantum multiverse, which, paradoxically, [has] less support among physicists than the cosmological kind.

4:32 **Robert Lawrence Kuhn** Now in your multiverse, the different universes, if we call them that, actually do interact, which is how you [generated] it in the first place.

4:41 **David Deutsch** Yes, they are in constant, intimate interaction. Although the experiments that can only be explained in terms of multiverse are very hard to arrange and they require very subtle laboratory techniques, in fact we know from the theory that practically all of our everyday experience is conditioned by multi-universe interactions. For example, matter couldn't be solid if each atom only took one path. What keeps it solid is the interactions between different instances of the same atom in different universes. That is what makes for rigidity. It's also for what makes permanent magnets, it's also the reason that we can have amplification of signals, it's the reason that we can see stars.

5:32 **Robert Lawrence Kuhn** Now clearly that is not a common accepted view.

5:36 **David Deutsch** I think all physicists who look at these phenomena would agree that quantum theory is needed to explain those phenomena. But they would not express quantum theory in terms of parallel universes.

5:48 **Robert Lawrence Kuhn** For example, the solidity of matter could be explained by the exclusion principle, because things just can't occupy the same spot, so it seems rigid when in fact it's mostly space.

6:02 **David Deutsch** Yes, so that's a good example. So if we take the exclusion principle and express it in its true quantum theoretic terms, it says that a sum of certain terms of a vast number, exponentially vast number of terms, has to equal some other exponentially vast number of terms. And that is saying that each one of these things, if it represents reality, represents a different history of the atoms, and they are affecting each other.

6:31 **Robert Lawrence Kuhn** When you're saying a different history, some people talk in quantum mechanics as though these are possible histories, but they don't have a basis in reality. You are claiming that they have a basis in reality?

6:42 **David Deutsch** Yes.

6:42 **Robert Lawrence Kuhn** I mean, this is a dumb question, but how many are there?

6:49 **David Deutsch** There are vast numbers.

6:49 **Robert Lawrence Kuhn** Oh, it's not good enough, vast. I want a number.

6:52 **David Deutsch** So, for example, if we're talking about the number of different histories that's happening in, let's say, a cubic meter of air, then we're talking in terms of ten to the power of ten to the 23 different histories are all happening at the same time.

7:08 **Robert Lawrence Kuhn** So that's ten to the power of...

7:13 **David Deutsch** Of number one with 23 zeros.

7:15 **Robert Lawrence Kuhn** And that billion, billion, billion, something like that.

7:18 **David Deutsch** Yes.

7:19 **Robert Lawrence Kuhn** And so that's the number of different histories of this one meter of air. And of course, we have a gigantic universe. Now, those histories, when do they occur? From the beginning of time or right now? I mean, is it sequentially or in parallel?

7:37 **David Deutsch** One has to understand it in this way. But first of all, yes, they have existed since the Big Bang. There was an older version of the quantum multiverse in which they only appeared as branches, but that has now been abandoned and in fact we now think of it in terms of happening for all time. But remember that there's a lot more in the multiverse than just universes, and in fact universes are an emergent property of the multiverse. So some types of phenomena like human thought and the outcomes of scientific experiments are expressible in terms of universes, but a lot of other phenomena such as quantum computers or the interior of molecules are not, and for those it's the multiverse rather than individual universes.

CAN SCIENCE PROVIDE ULTIMATE ANSWERS?

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - Can Science Provide Ultimate Answers?
Description	If we seek answers to ultimate questions of human existence, can science provide them? In other words, if science is unable to know something, is that something forever unknowable? Or are there ways of knowing beyond science? If so, why would we trust them?
Link	https://www.youtube.com/watch?v=j65pZMkQv48
Ideas	• Although science cannot solve every problem, reason can. And the quest for good explanations reaches far beyond science into all branches of philosophy. • The theory that the only kinds of good explanations are scientific ones is itself not a part of science and, therefore, rules itself out. • No scientific theory can ever predict the future growth of knowledge.
Topics	bad philosophy • empiricism • good explanations • good philosophy • scientism • testability • the limits of science

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Robert Lawrence Kuhn** David, you are a great espouser of the power of good explanations, the scientific method to make progress. Indeed, you're an optimist, very strongly in this regard. Let me turn it around. What are the limits of science?

0:15 **David Deutsch** One of the most important limits of science is that it isn't philosophy. Science only deals with the physical world and [discovers] regularities in the physical world and also [the] means of controlling the physical world. So that's one limitation, and what we call scientism is the purported application of science to problems that are really philosophical. Such as the question of whether animals really feel pain or not. We can tell whether animals' nerves are excited and whether their brains react to that, but whether an animal feels pain in the sense that humans do or merely reacts in the sense that a robot does—that is ultimately a matter of philosophy because it's only philosophy that can determine the criterion for science to use when trying to distinguish between those cases. So that's a limit of science. Trying to reach into philosophy is scientism.

1:18 **Robert Lawrence Kuhn** Now, the history of the progress of science has been one of expanding its boundaries. So we can't say for sure, even today, where those boundaries are because as far as we know it has been constantly expanding. And some would say that ultimately it can expand, not necessarily to answer every possible question of existence, but to answer every question that can be answered. And then there would be no room for philosophy or theology or whatever else.

1:50 **David Deutsch** We can't predict the future growth of knowledge. That, by the way, is another limitation of science. But I think it's much more likely that the thing that is omnipotent, the thing that can reach to everywhere and solve every problem, is not science narrowly conceived, but reason. The quest for good explanations reaches far beyond science into all these philosophical areas.

2:16 **Robert Lawrence Kuhn** So you differentiate good explanation from science. It's not a one-to-one relationship.

2:21 **David Deutsch** No, science is a special case of good explanations. It's good explanations applied to questions about the physical world.

2:28 **Robert Lawrence Kuhn** Some would say that the only kinds of good explanations are scientific ones.

2:33 **David Deutsch** Yes, well, I would say to those people that that theory is not part of science and therefore it rules itself out.

2:43 **Robert Lawrence Kuhn** Okay, now if we look at the concept of explanation, there are philosophers of science who are empiricists in which they say that the claim that anything is really real is impossible and extends beyond the human capacity. We can say there are regularities, there are observations, the observations always occur, all the things that you can literally predict. But to go that step to say, "It's an explanation," is a step too far.

3:14 **David Deutsch** Yes, by the way, this step [of] this trope of saying that science can only deal with predictions but not with understanding what reality is like is usually part of a piece of bad philosophy that's trying to rule out a piece of real science, as it happens in quantum theory when people try to say that parallel universes aren't real because we can't directly see them, we can only see their results. But the thing is, it doesn't work as a foundation for science, that theory. Consider, for example, the theory that dinosaurs existed. Now, nobody will ever see a dinosaur, as the Creationists never tire of pointing out. Nobody has ever seen one or will ever see one, at least not the ones that we claim existed in the past. All we see are fossils. So this empiricism would say, "Science can't make any claim about dinosaurs, it can only make a claim about fossils. This fossil will be found in a stratum with this fossil but in a different stratum from this other fossil." Now, this drains science of its entire purpose, which is to understand reality. Nobody would be interested in fossils if they were just patterns in stones. There are plenty of other patterns in stones, and some of them are more interesting than the fossils and certainly easier to come by.

4:46 **Robert Lawrence Kuhn** Well, the empiricists would say, "I am protecting science, I am protecting it from doing things that are irrational. Maybe those bones will be taken as an example and bad theories made of those, and so I'm going to keep science very constrained on the track of truth and not let it bloat beyond it."

5:09 **David Deutsch** This is why we have a criterion for what is or isn't scientific, namely testability. So the scope of science keeps growing as we find ways of making testable theories about things where previously we couldn't. A prime example is cosmology, which, if you look in an old dictionary, you will see cosmology listed as a branch of philosophy. But if you look in a modern dictionary, it's listed as a branch of physics. That's an example of the totalitarian character of physics, that it tends to envelop everything else. At any one moment, we can tell exactly where the limit of science is using the criterion of testability, testable theories. We have testable theories of dinosaurs, so it's legitimate to talk about dinosaurs as they were hundreds of millions of years ago as well as they are now in the form of fossils. The criterion of good explanation, that an explanation should be hard to vary, implies that criterion for good science, but it also tells us what is good philosophy, and that is vital. In the case of empiricism, it tells us that that is bad philosophy because you could rule out anything that was real by that criterion. In fact, the sense impressions that seemed like a good basis for science at the time when empiricism was invented a few hundred years ago turn out to be highly complex things which are not observed.

6:41 **Robert Lawrence Kuhn** So it's self-defeating.

6:43 **David Deutsch** Self-defeating, yes.

6:45 **Robert Lawrence Kuhn** So as you look at the flow of science from the past and into the future, would you say that ultimately there are limitations, but we can't know where they are today?

6:56 **David Deutsch** That's exactly right. Science has limitations. Reason, however, does not. There will be ultimate limits of science beyond which progress will only ever be made with philosophy, broadly speaking. We don't know what those are. They are almost certainly beyond the current limits of science.

WHY IS THE QUANTUM SO STRANGE?

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - Why is the Quantum so Strange?
Description	To know reality, one must confront the quantum. It is how our world works at the deepest level. What's the quantum?
Link	https://www.youtube.com/watch?v=mNP5w4n9sFU

Ideas

- The set of a universal computer's possible motions—that is, the set of all possible programs that could be programmed into it—is in one-to-one correspondence with the set of all possible motions of anything.
- A theory of computation within any laws of physics is a theory of how you can use physical objects to represent abstract objects. The theory of quantum computation is the theory of what kinds of information processing our deepest theory in physics, quantum theory, allows and forbids.
- A comparatively weak quantum computer could perform more computations simultaneously than could be performed by the entire visible universe if it was all made into (classical) computers.

Topics abstract objects and physical objects • Babbage • computation as a theory of physics • Eugene Wigner • laws of nature • mathematical proofs • quantum computation • quantum computers compared with classical computers • Turing • universality of computation • unreasonable effectiveness of mathematics

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Robert Lawrence Kuhn** David, quantum computing is becoming quite important in the world, to no small degree based on your personal contributions. I'm not so interested [in] the applications, can we factor large numbers and steal money through bank transactions or catch spies, but what it is fundamentally, and what it possibly can tell us about the nature of reality?

0:25 **David Deutsch** That's really why I'm interested in it as well. I came to this from a physics point of view, not a computation point of view. One of the things that really attracts me about the theory of quantum computation is what it tells us about what kind of thing a law of physics is. It's been a mystery to philosophers and physicists for decades, what I think Eugene Wigner called the unreasonable effectiveness of mathematics in the natural sciences. Especially when we realize that the set of computable functions, which are familiar to us, made of things like addition and multiplication and so on, from a mathematician's point of view, they form an infinitesimally tiny subset of the set of all possible mathematical relationships. And yet, physics is made entirely out of those, and if it weren't, we wouldn't be able to know any physics.

So when it became more and more obvious that computation is built into the laws of physics at a fundamental level, a lot of people immediately jumped to the conclusion, "Oh well, the reason that mathematics is useful in the physical sciences is that the world is a computer, and we are just programs running in that computer," or something like that. Or, "We're just a simulation running in a computer." Now it seems to me that that misses the whole point of the lesson of the universality of computation for physics, because it requires a notion of what is or isn't computable that is outside the physical world, so that it was set by God or something to be a certain set, and that's why our universe only instantiates that set of mathematical relationships. Well, that doesn't solve the problem. You may as well have said that God set up just our universe with those relationships.

2:31 **Robert Lawrence Kuhn** Eliminate the middleman.

2:32 **David Deutsch** Yes, remove the middleman. I think the real, important lesson of the universality of computation, as revealed by quantum computers to be part of physics, is that universal computers can be built within the universe. That is really the amazing thing, because however the universe was, you could imagine some kind of supercomputer with unknown mathematics that simulated it. But the amazing thing about our universe is that you can make an object, a computer that can simulate any physical process—that's what universality is—and this object, the set of all its possible motions, that is, the set of all possible programs that could be programmed into it, is in one-one correspondence with the set of all possible motions of anything. And that is telling us something about the universe from the inside. It's telling us something about what laws of nature actually are.

3:41 **Robert Lawrence Kuhn** Okay, now how does the quantum part help us?

3:46 **David Deutsch** When the theory of computation was first discovered by Babbage and then developed by Alan Turing during the 1930s, it wasn't realized that this was a branch of physics at all. It was invented as a branch of mathematics to study mathematical proofs. And the theory was built up from a conjecture that a certain type of abstract object, the Turing machine, could represent all things that could be computations. And then historically what happened after that is that people began to worry that the physical world might not be able to instantiate these operations perfectly and that, therefore, the real world might be a weaker kind of computer than the Turing computer, that it might be an idealization.

When we studied this more carefully, and this is where quantum computers begin to come in, we found that not only can a universal computer exist physically, but it's more powerful than a Turing machine. And what the mathematicians were doing unconsciously is that when they invented these abstract objects, they were applying their intuition about physical objects. They didn't know that that's what they were doing. And because they were applying their intuition about physical objects, they got it wrong. They thought about computing, making marks of squares of paper, and then, as Feynman remarked, "They thought they understood paper." But in fact, paper, like everything else, obeys quantum mechanics, and the real computation in the world is quantum computation. The theory of computation is the theory of quantum computation, and that is a theory of physics. So that means that the theory of computation is irretrievably within physics because of the quantum theory of computation.

5:49 **Robert Lawrence Kuhn** Now what is, briefly, the quantum theory of computation? How does that work? How is computation and quantum theory, quantum mechanics, integrated into a quantum theory of computation?

6:03 **David Deutsch** A theory of computation within any laws of physics is the theory of how you can use physical objects to represent abstract objects. So you want to represent the integers 1, 2, 3, and you can use physical objects like fingers to say, "That will be 1, that's called 2, that's called 3," and so on. And computers are ways of instantiating abstract objects and their relationships in physical objects and their motion. Now what happens with quantum computers is that we simply take the deepest physical theory we have, quantum theory, and we say, "What kind of information processing does quantum theory in general allow and what does it not allow?" And that's the theory of quantum computation.

6:50 **Robert Lawrence Kuhn** And when you do that, what do you find compared with a classical computer when you make this quantum computer?

6:58 **David Deutsch** You find a number of similarities, and we find the reasons why the Turing theory worked as well as it did, and then you find a number of dramatic differences between the quantum computers and classical computers. The one that's got the most attention is that, for certain types of calculation, a quantum computer can perform it exponentially faster than any classical computer. People haven't built quantum computers yet, but we hope that they soon will. And when a quantum computer is built, a small quantum computer with a few thousand qubits, that's the quantum analog of bits...

7:42 **Robert Lawrence Kuhn** Compared to the billions of bits in our normal desktop computers or laptops.

7:45 **David Deutsch** Yes, or even our mobile phones. In other words, a very, very comparatively weak quantum computer could perform more computations simultaneously than could be performed by the entire visible universe if it was all made into computers. In fact, when I say more, that's an understatement. Exponentially more than that. But only for certain types of computation, and that's a token of the fact that the whole notion of computation is different in quantum computers. It's not that, like with all classical computers, you can say that one computer is ten times as fast as the other. With quantum computers, they are vastly faster than classical computers for some computations, and the same for others. And interestingly, they're not slower for any computations because a quantum computer, among its abilities, is to simulate a classical computer.

WHAT IS TRUTH?

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - What is Truth?
Description	Defining 'truth' is an ancient question that in the age of science should find resolution and agreement. But this is not so. Even today, #truth remains elusive. Can truth be objective or must it always be relative?
Link	https://www.youtube.com/watch?v=3eEffbjzNwE

Ideas

- Good explanations are extremely hard to come by, and this is what the growth of knowledge is actually about. A good explanation is one that is hard to vary while still explaining what it purports to explain.
- If we find a good explanation at a higher level of emergence, then it's irrational to reject it just because it doesn't have a form that we prefer.
- The idea that all progress comes from the quest for good explanations distinguishes between ideas that have a chance of making progress and ideas that have no chance of making progress in every field—science, political philosophy, moral philosophy, aesthetics, etc.

Topics bad philosophy • emergence • fundamental law • good explanations vs. bad explanations • levels of explanation • logical positivism • modes of explanation • Newton's theory vs. Einstein's theory • principle of testability • progress • reductionism • statistical analysis • the growth of knowledge

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Robert Lawrence Kuhn** David, you have this concept of explanation, which is a normal-sounding word, but you use it to really probe not just the fundamental aspects of reality, but where humanity can go in the future. It's a very powerful part of your philosophy. So I'd like to understand what you mean by 'explanation.'

0:24 **David Deutsch** An explanation is a statement of what is there in reality, and how it works and why, basically. But the important distinction is between a good explanation and a bad explanation, because explanations are two a penny, but good explanations are extremely hard to come by, and this is what the growth of knowledge is actually about. So a good explanation is one that is hard to vary while still explaining what it purports to explain. Shall I give an example? Suppose you're watching a conjuring trick, and you're trying to explain what's happening. Now, an example of a bad explanation would be, "Well, it's actually magic." And the reason that that's a bad explanation is that you could apply that same explanation to absolutely anything, including to the conjuring trick happening a different way, or to a conjuring trick not happening, and so on. So that claim about reality, that it really is magic, is not actually an explanation, or, if you like, it's a bad explanation.

Another example of a bad explanation, just to show you that a bad explanation doesn't necessarily have to be false, it just may be completely inadequate, is to say, "Well, the conjurer did something." So that may be enlightening for a person who believes in magic, to tell them that in fact it was the conjurer that did it, but it doesn't explain the trick. And if we take by analogy with the laws of physics, [in] trying to explain things in the natural world, we could say, "What is the origin of species? What is the origin of adaptations in the biological world?" You could say, "Well, it's just caused by atoms." Now that's true enough, but it doesn't explain. The explanation is Darwin's theory of evolution, or rather the modern neo-Darwinist theory of evolution.

2:40 **Robert Lawrence Kuhn** So you've now differentiated good explanations from bad explanations. How does this apply? For example, explanations in science normally have a reductionist approach, which says that in order to explain what's on higher levels, like we're human beings, you have to understand systematic organs, and to understand the organs you have to understand cellular structure, and [from] cellular structure, biochemistry, and then physical chemistry, and down to physics, and fundamental physics, and now you have a complete explanation.

3:10 **David Deutsch** Yes. This reductionism is a prejudice. It's historically understandable because the physical sciences, especially physics, were the ones that developed fastest. And it so happens that the best explanations in physics have been, at any rate, from the ground up, from space and time, elementary particles, that kind of thing. But it's never been the case, even within physics, let alone in other sciences, that all good explanations are reductionist. And in fact, my basic principle, if you like, that we should be looking for good explanations, which I think is the foundation of scientific rationality, implies that we must not have that prejudice, because if we do find an explanation that's on a higher level of emergence, say, and we find a fundamental law at the higher level of emergence, and it's a good explanation, then it's simply irrational to reject it just because it doesn't have the form which, historically, we have been taught is the one we should pursue.

4:16 **Robert Lawrence Kuhn** So by really understanding the deep power of explanation, you become more open to different modes of explanation?

4:24 **David Deutsch** That is exactly right. And I think with deep explanations, it's nearly always the case that when somebody finds a new and much deeper theory, it's not only a better explanation that they've found, it's also a better mode of explanation. So for example, in physics, Einstein's explanation of gravity in terms of curved spacetime was not just a new explanation of gravity, that would have been something like Newton's laws, but instead of an inverse square law, an inverse 2.003 or something law. It's not like that. It's a different kind of explanation. It's saying that space and time, which in Newton's theory are immutable background entities that aren't part of the theory, become, in Einstein's theory, dynamical objects which buck and weave and explain all sorts of things apart from just the motion of planets.

5:26 **Robert Lawrence Kuhn** What I like about your approach to bad explanations is that they're not only false, but they disturb your ability to even make progress or to find out what are good explanations.

5:38 **David Deutsch** Actually that definition is what I call 'bad philosophy.'

5:42 **Robert Lawrence Kuhn** Okay.

5:42 **David Deutsch** Bad philosophy…

5:43 **Robert Lawrence Kuhn** Is a subset of bad explanations.

5:44 **David Deutsch** Yes, it is a subset of bad explanation.

5:47 **Robert Lawrence Kuhn** So tell me about bad philosophy.

5:48 **David Deutsch** False philosophy is not harmful. In fact, error is the standard state of human knowledge. We can expect to find error everywhere, including in the theories that we most cherish as true. But there has grown up, especially since the Enlightenment, ironically, since good explanations have begun to take over, bad explanations have become worse and bad philosophy has dominated the field of philosophy for many decades. Bad philosophy is philosophy whose effect is to close off the growth of knowledge in that field. So it's the kind of thing that says not just so-and-so is true when in fact it is false, but you mustn't think about so-and-so, or it's bad to investigate so-and-so.

6:44 **Robert Lawrence Kuhn** What's an example? Logical positivism?

6:46 **David Deutsch** Logical positivism is a prime example of a bad philosophy.

6:51 **Robert Lawrence Kuhn** Which restricts your ability to even address questions as meaningless because it's not [in the] sense data or logic or something like that.

6:58 **David Deutsch** Exactly right. So it's saying that trying to understand what the physics is of unobserved objects is unscientific, according to positivism. Now that means really that it's trying to reduce us to an anthropocentric worldview, rather like the medieval worldview, because it's saying that the only things that are worthy of study are human experiences. But of course human experiences are themselves to be understood in terms of unexperienced things like neurons. So the whole philosophy collapses, and in addition it declares itself to be meaningless because this distinction that it draws applies to itself as well and rules out positivism as a worthy subject of study.

7:50 **Robert Lawrence Kuhn** What are other examples of bad philosophy?

7:53 **David Deutsch** The ones that are closest to my mind are the ones that have impinged on physics. So logical positivism was one example of that. But in recent times, statistical analysis of experimental results has started to use terminology that assumes that certain things will never be worth studying. So, for example, the very term 'explanation' has come to mean a mathematical formula. They say a mathematical formula explains the results. But since the results are anthropocentric and they are not reality, they're just a tiny sliver of reality through which we are trying to understand the unobserved reality, this idea that a formula is an explanation prevents real explanations from being discovered. They are ruled illegitimate.

9:02 **Robert Lawrence Kuhn** So it's almost, not just circular reasoning, but it confines you within an area that you're unable to get out of.

9:10 **David Deutsch** Yes.

9:11 **Robert Lawrence Kuhn** Okay, so what is the antidote?

9:15 **David Deutsch** I think that all progress, historically and today, comes from the quest for good explanations, that is, explanations that are hard to vary while still accounting for what they purport to account for. One of the reasons I like this principle is that not only does it explain what the criterion for success is in science, where it leads to things like the principle of testability of theories because a test constrains the explanation so that it's hard to vary, but it also applies outside physics in philosophy, in epistemology, in metaphysics, and so on. The same thing applies and even beyond that in political philosophy, moral philosophy, and aesthetics. The same principle applies everywhere and draws a distinction between ideas that have a chance of making progress and ideas that have no chance of making progress.

IS THE COSMOS A COMPUTER?

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - Is the Cosmos a Computer?
Description	That the cosmos is a computer sounds like a modern metaphor, a way of explaining how things work. But some make a bolder claim: that the cosmos is in reality a computer, not just as metaphor.
Link	https://www.youtube.com/watch?v=UohR3OXzXA8
Ideas	• The existence of universal computers (or their possibility) explains the unreasonable effectiveness of mathematics in the natural sciences. • Explaining reality in terms of qubits, bits, and computations is not an advance in explanation over explaining reality in terms of space, time, atoms, and forces. Still, information and computation are fundamental features of the physical world in the sense that there is a law of nature that says that the universe is computable. • Universal computers are inherently emergent objects. You could create a computer at the subatomic level, but then it could not be universal and, therefore, not fundamental.
Topics	fundamentality of information • laws of nature • the reductionist mistake • universal computer

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Robert Lawrence Kuhn** David, there seems to be a fad in physics today that information is the most fundamental thing in the universe. It's not just that information describes matter and energy and fields, but at the bottom of matter is information. Information is the most fundamental. And as a consequence, the universe itself is not a metaphor like a computer, but in fact is a computer, and that's what reality is.

0:29 **David Deutsch** Yes, I'm opposed to that view. I think it misses the point, it misses the lesson of what is really important about the link between computation and physics. Computation, and with it information, is indeed very fundamental, but it's not the most important thing. It's not the thing that the universe is made of, as it were. So there are variants of this theory, like the universe is a computer, or that the universe is a program running in God's computer, or something like that. It seems to me that those theories all miss the point about the universality of computation, which is the way that computation links with physics, through the existence of a universal computer, a computer that can compute anything that can be computed, and can therefore simulate any physical object. If this computer were outside the universe, it wouldn't be very remarkable. You can always imagine some kind of computer with some kind of way of operating that would simulate any laws of physics, no matter what they were. And so you lose the fact that our actual laws of physics are intimately connected with computation.

What is the connection? It's not that there's a computer outside the universe, it's that we can make universal computers inside the universe. That is a token of the computability of the laws of nature. It's the reason for the unreasonable effectiveness of mathematics in the natural sciences. It's the reason for the existence of life and the possibility of science. Those things are explained by the existence of computers in the universe, but wouldn't be explained if the universe was in a computer.

2:22 **Robert Lawrence Kuhn** If that is true, how do you account for this growing sense that at the bottom, the bedrock of reality is information? That it—stuff—comes from bit—information?

2:39 **David Deutsch** It's perhaps a natural mistake to make, given the fundamentalness of computation. Perhaps the reason for the mistake, now that I come to think of it, is that we are accustomed to explaining things in reductionist terms. So the new idea is, instead of explaining things in terms of space, time, atoms, and forces, we explain it in terms of qubits, or bits, and computations. But that is not an advance in explanation, for the reason that I just said. And the real connection is the computability of the laws from within the universe.

3:29 **Robert Lawrence Kuhn** And so that makes information part of the process of the universe, but not sitting at the fundamental nature of the universe?

3:38 **David Deutsch** Its existence is indeed fundamental, but it's not fundamental in the reductionist sense. It's fundamental in the sense that there is a law of nature that the universe is computable, or that a universal computer exists.

3:52 **Robert Lawrence Kuhn** And would this computation, this information, work on various levels in a hierarchical sense, or would it only exist on the most fundamental level, as you understand information?

4:05 **David Deutsch** Yes. Well, computers are emergent objects. There is no such thing as a computer at the subatomic level. Or if you did make a computer at the subatomic level, it wouldn't be universal, and therefore wouldn't be fundamental. The fundamental computers are the ones that are universal, and they are the ones that are quite big. They're like the computers that we actually have and use in everyday life.

4:30 **Robert Lawrence Kuhn** But information could operate on the most fundamental level.

4:33 **David Deutsch** It could and does, but that, as I have argued, can't underlie…that's just an alternative way of talking about atoms.

4:42 **Robert Lawrence Kuhn** Okay, but this concept of information, even if it's not at the most fundamental level, when it exists, as you see it, does it exist at each level of the hierarchy of the laws of nature, or is it only existing at the level of physics?

5:01 **David Deutsch** Yes, good question. Information exists at every level, including at the fundamental level of atoms. Whether an atom is there or not is a fundamental thing. But at the level of laws of nature, information comes in at a particular level of explanation, namely the level at which there are computers, and the level at which there are people thinking about stuff. That level of explanation.

5:29 **Robert Lawrence Kuhn** How then does that way of thinking allow us to understand reality better?

5:35 **David Deutsch** Because almost all ways that the laws of physics could be do not have the property that a universal computer could exist.

WHICH LAWS OF NATURE ARE FUNDAMENTAL?

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - Which Laws of Nature are Fundamental?
Description	Why is there a world that works so well? How does the cosmos generate diversity and opportunity?
Link	https://www.youtube.com/watch?v=2BLo2SdmjLI

Ideas

- In the 1970s, physicist Brandon Carter discovered that if the electron's charge had been only a few percent different, then there would be no complex chemistry and, therefore, no opportunity for life to evolve. The solution to this fine-tuning problem may be an as-yet undiscovered law of physics about emergent properties that implies that people must exist.
- Constructor theory is the generalization of the theory of computation to the rest of physics. While the theory of computation is about which computations can and can't be brought about, constructor theory is about which transformations more generally can and can't be brought about.
- The transformations that people can achieve in real life are precisely the ones that are not forbidden by the laws of physics. For example, if there isn't a law of physics that says you can't live to be 500, then living to be 500 is a soluble problem. It's just a matter of knowing how to achieve it.

Topics

anthropic self-selection • Brandon Carter • constructor theory • Copernican principle • fine-tuning problem • fundamental laws • laws about emergent properties • optimism • Richard Feynman • the fabric of reality

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Robert Lawrence Kuhn** David, the laws of physics seem incredible in that they are perceptible to us, we can manipulate them, we can use them for predictions. What does that begin to tell us in terms of their fundamental nature? And how can we begin to look at the laws of physics and see what the nature of reality is?

0:23 **David Deutsch** It's certainly the case, and I think this is now uncontroversial, that if the laws of physics were very slightly different in almost any way, there could be no life in the universe, no complex chemistry, and no thinking people, and therefore no one who knows the laws of nature. So they are somehow almost infinitely special in that they allow themselves to be, as you said, not just known, but also used, and they were used before humans even existed to create life and then for the human species to evolve. Now, that has been for several decades an unsolved problem at the foundations of physics, why that is so, called the fine-tuning problem. And it began in a serious way… People began to investigate this in the 1970s. The physicist Brandon Carter, who was investigating the evolution of stars, found that if the charge on the electron had been only a few percent different, either larger or smaller, then there would be no complex chemistry and no opportunity for life to evolve. So the standard take on this is that this is evidence that the laws of physics as we see them are not the only ones that are instantiated in physical reality. It's rather like the argument, you know, you win the lottery and you say, "Why me?" It seems very strange that the lottery should have picked out you. And the solution to that is, you'll realize that that's not such a strange thing if [you] realize that a million people entered the lottery, and one of them had to win.

2:23 **Robert Lawrence Kuhn** Or if you hit a golf ball and it lands on a blade of grass, you say, "What are the odds that it's on that blade of grass?" You know, one in however many blades of grass there are on the field.

2:34 **David Deutsch** But the thing that makes the fine-tuning problem more mysterious than just any old random number like a lottery or a blade of grass is that the particular blade of grass that it landed on seems to have a purpose, seems to be tuned, as they call it, for our existence. And this seems to violate one of the first things that was realized at the beginning of modern science, which is that humans are not especially distinguished by the laws of physics as the center of the universe or as the purpose of the universe or anything like that, but that everything about us is explained by laws that don't particularly refer to us.

3:17 **Robert Lawrence Kuhn** The Copernican principle.

3:18 **David Deutsch** Yes, yes, that's right.

3:20 **Robert Lawrence Kuhn** So the explanations that have been given are that—and they're radically different, and these are pretty much the only two explanations—is that in one way we have been designed to be special by some Creator, God, that some people would like, or some superintelligent species in which we're simulated, some sort of a creative process, maybe not necessarily a traditional god, some sort of creative process. The other extreme are multiple universes in a cosmological sense, which each one of these multiple universes, an infinite number perhaps, picks out different laws of physics so that, in the process of this randomized approach, one or more would give rise to us, and we're in that universe, so it's the only one we're in, so we're asking the question, "Why are we special?"

4:09 **David Deutsch** Yeah.

4:09 **Robert Lawrence Kuhn** One of those two. That's what we're given. Do you like either one of those?

4:13 **David Deutsch** No. I think both of those are incapable of solving the problem. The first one, the idea that the laws of physics were designed by someone or something, simply raises the question that that thing also has to be fine-tuned. It also has the very properties that we're wondering about the origin of in ourselves.

4:35 **Robert Lawrence Kuhn** Kicks the problem up a level.

4:37 **David Deutsch** Yes, without making it any better. It's okay to kick the problem up a level if you then have an easier problem, but if you have the very same problem, then that's an infinite regress.

4:47 **Robert Lawrence Kuhn** Or it might be a harder problem if that's a nonphysical thing.

4:50 **David Deutsch** Could be an even harder problem, in which case it's worse than an infinite regress. Now, the other idea, which is the one that is greatly favored by cosmologists currently, I'm not entirely sure why, but it has become the prevailing theory in cosmology, is this idea that there's an ensemble, a vast set of different universes. Now the trouble with that, as was pointed out by Richard Feynman many decades ago, is that if the only explanation of why the laws of physics seem to favor us is that if we weren't here, we wouldn't be asking, the overwhelming majority of universes in which someone is asking, they are only just asking. That is, the universe is only just good enough. There are many more universes where, for example, this room and its contents have just sprung into existence and will disappear immediately afterwards.

5:52 **Robert Lawrence Kuhn** A fluctuation.

5:54 **David Deutsch** Of just a fluctuation. And this idea that the universe could be just one in an ensemble suffers from the fatal flaw that most such universes that have the property of containing us only just have it, and we're about to die because a sphere of heat is coming in at the speed of light and will extinguish us in the next picosecond. So that means that some principle, other than just anthropic self-selection, has to be responsible for the fine-tuning, and it can't be design because that just kicks the problem upstairs.

6:43 **Robert Lawrence Kuhn** It sounds like there's no solution because I don't got one. I'm waiting how you can solve this.

6:47 **David Deutsch** I don't pretend to have a solution, but I think I have an argument why there can be a solution apart from those two. If the solution isn't either of those two, then the solution is a law of physics. It's a law of physics that applies in our universe or perhaps in our universe and a trillion others. But just having, as I said, just having multiple universes doesn't solve the problem. They would have to be multiple universes that are tuned so that most things in them don't only just exist. I think the key is that the laws of physics as we currently conceive them are based on atoms and working out everything that happens from a microscopic level. But if we admit into fundamental physics laws about emergent properties such as computation, one of those may imply that we exist without being anthropocentric.

7:50 **Robert Lawrence Kuhn** David, as we consider the laws of nature, we always try to find those which are the most fundamental. And physicists would have us go deeper and deeper in a reductionist sense to try to find those laws. How do you look at even approaching the problem?

8:11 **David Deutsch** What I take to be a fundamental law is one that is implicated in many other explanations. And the most fundamental laws in physics happen to be reductionist laws of quantum theory and the theory of relativity. Although there are non-reductionist laws like the second law of thermodynamics, even in physics. But there are other laws. The principle of evolution, for example, which says that adaptive complexity can only arise through variation and selection, is a rigid law of nature and yet is intrinsically emergent. So that's another law. The laws of epistemology that say that knowledge is acquired by conjecture and criticism, that's another rigid law.

8:57 **Robert Lawrence Kuhn** So now you've given three radically different kinds of laws, from fundamental physics, to biology of species, to approach to knowledge, that you're saying are all fundamental but are radically different even in their categories.

9:15 **David Deutsch** Yes, they are all fundamental in that they are needed to explain many things, and we can't explain everything in terms of just one of those strands.

9:24 **Robert Lawrence Kuhn** And therefore, to you, explanation is an organizing principle that can unite those.

9:31 **David Deutsch** That's right. One of the things that looking at it this way helps with is that we can see that laws at different levels of emergence actually mesh together into what I call the fabric of reality, into a sort of unified worldview, which we can then extend. One of the things I'm trying to work on now is extending the theory of computation into the theory of not just what can and can't be done with abstract objects, but the theory of what can and can't be done with any [objects], which is a way of looking at physics in the manner of the quantum theory of computation. And remarkably, that connects not only physics and computation, but it also has all sorts of philosophical implications, such as optimism comes out of that theory.

10:26 **Robert Lawrence Kuhn** Well, we certainly need some optimism, but I'm at a loss to see how we can get optimism from where we are, so walk me through. What do you call this theory?

10:36 **David Deutsch** It's called constructor theory. It's the generalization of the theory of computation to the rest of physics. And the way it's linked to optimism is very simple. If you imagine the set of all transformations. We want to transform the world into a better world, let's say. Now, some of those transformations are permitted and some are not permitted by the laws of physics. So the question is: Which ones of them can we actually achieve in real life? And the answer to that must be, according to constructor theory, that the ones that we can achieve in real life are precisely the ones that are not forbidden by the laws of physics. So if the laws of physics say we can't travel faster than the speed of light, then we never shall. But if there isn't a law of physics that says you can't live to be 500, then living to be 500 is a soluble problem. It's just a matter of knowing how.

11:36 **Robert Lawrence Kuhn** So what are the limitations of physical laws that will give us those ultimate constraints? Because anything within those constraints is ultimately achievable.

11:47 **David Deutsch** That's right. The laws of physics are not actually very onerous in regard to achieving what humans want to achieve. Even traveling to another galaxy, although you can't do it in the time, fortunately relativity means that your time will slow down if you travel very fast. So if you really wanted to travel to another galaxy in your lifetime and you had the right technology, you could do so subjectively. So it's not very onerous. The things that we are accustomed to calling evils, even the ones that are deemed to be inevitable evils like death, are actually just a matter of technology to solve.

12:30 **Robert Lawrence Kuhn** So you look very optimistically in terms of what technology can achieve.

12:35 **David Deutsch** Yes. And this, as I said, follows from very fundamental considerations within physics. The thing is, if there were a thing that we can't achieve no matter what knowledge we bring to bear, let's say it was living to 500 or something, suppose that there's no law of physics that we can't, but we still couldn't achieve it, well then if we can't achieve that no matter what knowledge we bring to bear, then there is another law of physics that says that we can't do that. And that's a testable law.

A testable regularity in nature is a law of physics.

13:09 **Robert Lawrence Kuhn** So as we push forward, as we push knowledge forward, as you would like to say, infinitely forward, as we do this...

13:17 **David Deutsch** Or unlimited, yes.

13:18 **Robert Lawrence Kuhn** As we do this, we will either make progress or discover new laws of physics that constrain us, one or the other.

13:26 **David Deutsch** Exactly.

DO GENERAL PRINCIPLES GOVERN ALL SCIENCE?

EPISODE DETAILS

Date	2012
Interviewer	Robert Lawrence Kuhn
Source	YouTube
Show	Closer to Truth
Episode	David Deutsch - Do General Principles Govern All Science?
Description	Are there 'general principles' that encompass all sciences, which have explanatory strength from physics to biology? Could such general principles even explain actions and activities beyond the physical and biological sciences, such as in psychology, sociology and economics?
Link	https://www.youtube.com/watch?v=YD-capWJMyc
Ideas	• If you look in sufficiently fine detail at the boundaries between morality, aesthetics, epistemology, and science, you find they merge into each other and they can't be separated. • The distinction between better or worse exists objectively in aesthetics as it does in morality and in every other area of philosophy. • Laws of morality are not written in the language of mathematics, but we may improve upon them via rational analysis.

Topics	general systems theory • objective aesthetics • objective morality • objective truth • Popper's criterion in politics • testable theories • the principle of good explanation • unity of nature

Episode title and description reproduced verbatim from the original source.

Transcript

0:05 **Robert Lawrence Kuhn** David, we see principles at work in each field of science, in physics, in biology, and even in the social sciences, different ways of organizing observations. And some would say that some of these principles are very similar, so they try to build so-called general systems theory that take observations and laws from different parts of human knowledge and look for deep underlying principles that can be applied in each of these. Does that make sense?

0:39 **David Deutsch** Yes. I'm not sure that the existing approaches to general systems theory is the actual way of integrating all sciences, but I think the idea that all sciences are integrated by their principles at the fundamental level is correct and has to be correct. An obvious principle that unites all science is just the principle of testability, that the truth about nature takes the form of testable theories. I think that the principle of testability is a special case of a much more general principle, the principle of good explanation, a good explanation being one that is hard to vary while still accounting for what it purports to account for.

1:23 **Robert Lawrence Kuhn** Well, there are things that perhaps are good explanations that cannot be testable. What I like in music, I may like Mahler's Second Symphony and you may like Brahms' First Symphony as your favorite. Now those are real facts about the world, but they're certainly not testable in any way. But they might have a good explanation.

1:46 **David Deutsch** Exactly. So what characterizes science within the realm of human knowledge, is that science has testable theories and the truths about the physical world consist of testable theories. But this idea of a good explanation reaches beyond science into even, you mentioned aesthetics, even aesthetics. It's customary to say so-and-so is a matter of taste to mean there is no truth of the matter. But I think that cannot be so. I think there is a truth of the matter. It really is objectively true that, for example, Mozart produces better sounds, more aesthetic sounds than cavemen banging rocks together. And although we may not have a sophisticated enough knowledge of aesthetics, especially in explicit form, to know which is which, we know that it is there. The distinction between better or worse exists objectively in aesthetics as it does in morality and in every other area of philosophy.

2:52 **Robert Lawrence Kuhn** That's a fairly dramatic statement because to defend it by comparing Mozart to cavemen with their rocks sounds like it makes sense. But now if you compare Mozart to Beethoven or Mozart to Brahms, I don't think you can have an objective analysis.

3:11 **David Deutsch** What's happening there is that we do not know yet what the better way of analyzing these things is.

3:17 **Robert Lawrence Kuhn** But in that case, is that analyzable even in principle?

3:21 **David Deutsch** I think it must be, and for the following reason. You cannot separate these fields, science and aesthetics and so on, totally from each other. As Jacob Bronowski said, for example, "You can't do science, you can't make progress in science unless you also have certain moral values such as tolerance, respect for the truth," and so on. So these things are matters of moral philosophy, but they are essential to science as well. And therefore, they are essential to how the physical world is put together. So these different fields are only separated from each other for pragmatic reasons. If you look in sufficiently fine detail at the boundary between all these different fields of philosophy and between philosophy and science, you find they merge into each other and they can't be separated.

4:15 **Robert Lawrence Kuhn** So we have a number of ideas that we can classify as the principles that you feel really do work. Testability and good explanation. Are there any others that fundamentally can be used to unify the sciences?

4:30 **David Deutsch** I think that good explanation is the fundamental one, as far as is known at present. I don't believe that there's ever [an] absolute foundation to be found to knowledge, but I think the deepest thing we know at the moment is the principle of good explanation, which implies all sorts of things. About science, it implies the principle of testability. In politics, it implies Popper's criterion that institutions should be constructed in such a way that governments and policies can be removed without violence and so on.

5:08 **Robert Lawrence Kuhn** So basically you are saying that general systems theory is correct, but it's only correct if we have one general systems theory principle, and that's good explanation. And within that broad category, there are various subsets, including testability and science.

5:25 **David Deutsch** Yes. As far as we know.

5:27 **Robert Lawrence Kuhn** Now, an explanation would not have to have a quantitative comparison as a requirement.

5:35 **David Deutsch** That's right. Galileo said that "The laws of physics are written in the language of mathematics," but the laws of morality are not. And the laws of aesthetics are not. Probably aren't. We don't know much about the laws of aesthetics.

5:49 **Robert Lawrence Kuhn** And about human society, whether it's politics or sociology, some of that may be absolute and some of it may not, but even that which is not subject to quantitative analysis is subject to rational analysis, which is part of a good explanation.

6:07 **David Deutsch** Exactly. Rational analysis and objective truth, whether or not it's quantitative. The aspiration of general systems theory is definitely right. And in all these fields there is such a thing as objective truth to be found. And that is part of what will link them, but whether the actual ideas in general systems analysis are currently right, I doubt.

MICAH REDDING:

HUMANITY'S INFINITE REACH

About the interviewer: Micah Redding, Executive Director of the Christian Transhumanist Association, is a software developer and writer on the subject of human values and technology. He grew up as the son of a Church of Christ preacher, and is a fourth-generation graduate of a small Christian university. Micah currently lives in Nashville, Tennessee, where he has helped to organize vibrant inter-religious dialogue, leading conversations between Christians, atheists, Buddhists, Baha'is, and Muslims — and challenging Christians to think deeply about the meaning and significance of their faith. He spent several years promoting and working with a large annual charity walk, and has coordinated community events like cutting-edge art festivals and public conversations on technology and the future of the human race.

About the Christian Transhumanist Association: They believe that God's mission involves the transformation and renewal of creation including humanity, and that they are called by Christ to participate in that mission: working against illness, hunger, oppression, injustice, and death.

The Christian Transhumanist Association homepage:
https://www.christiantranshumanism.org/

EPISODE DETAILS

Date	July 28, 2017
Interviewer	Micah Redding
Show	The Christian Transhumanist Podcast
Episode	David Deutsch & Humanity's Infinite Reach
Description	David Deutsch explains the physics of humanity's profound place in the cosmos—and our potentially unlimited future.
Link	https://www.christiantranshumanism.org/podcast/35/
Ideas	• Karl Popper said that what is needed for progress is a tradition of criticism. Such a tradition stabilizes the changes that follow from criticism, and then you can have the growth of knowledge and progress. • There is one advantage that good has over evil, which is that the bad guys are wrong. The enemies of civilization are wrong, and therefore they, unlike the good guys, have an inherent interest in preventing their ideas from changing. And that's why they are always worse at generating knowledge than the good guys. If we try to rein in our own knowledge creation, then we are destroying our only advantage. • Personhood is a property of a program. It's a property of a computer program or a brain program.
Topics	AI vs. AGI • cosmology • dark energy • democracy • enemies of civilization • explanatory knowledge • genetic knowledge • institutions • jump to universality • knowledge as information that can be used to achieve physical transformations • Martin Rees • momentous dichotomy • Omega Point theory • personhood as a property of a program • philosophical significance of people • Popper's criterion for political institutions • Spaceship Earth • the Enlightenment • the quest for good explanations • traditions of criticism • unforeseeable problems • universality

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Micah Redding** Hey folks, Micah here. We're about to get started, but before we do, I just want to remind you that you can always get show notes for this and every other episode at christiantranshumanistpodcast.com. And while you're there, be sure to sign up for email updates so we can let you know when new shows are released, when new things happen in the Christian Transhumanist community, and most importantly so that we can connect you with other people just like you, exploring questions just like this. Thanks so much for listening, enjoy the show. I'm Micah Redding and I'm here with David Deutsch, who's an Oxford physicist, a pioneer in the field quantum computing, the founder of constructor theory, and the author of some of my favorite books, *The Fabric of Reality* and *The Beginning of Infinity*. Thank you so much David for joining me today.

0:53 **David Deutsch** You're welcome. It's nice to talk to you, and that's nice that you like the books.

0:58 **Micah Redding** I love them, and I've recommended them to so many people over the years. I was listening to your interview with Sam Harris, and he started out with an apology. It would be appropriate for me to apologize as well, because some of the stuff you deal with is so deep and so profound that it's just going to be impossible to do it justice. And I know some of my listeners will be people who are big fans of your work and some will be new to it, so to both of them, we won't be able to really do all of your thought the justice it deserves. But I was thinking that maybe the place that would be best for us to start is the question of humanity's place in the cosmos. Because there's this idea that humanity used to think of itself as the pinnacle of creation, the pinnacle of the universe, and then science came along and dethroned us and showed us that we're a mediocre, insignificant species on an insignificant planet in an insignificant galaxy. What's your take on that concept?

2:26 **David Deutsch** Well, as you know, I comprehensively reject that. I should say I conclude comprehensively that it's false. I don't take that as an axiomatic principle. It's just that it conflicts with what I think is our best knowledge today, both in the sense of physical effects and in the sense of theoretical significance.

I suppose the physical effects aspect is the easier one to explain. It's just that as we learn more and more about physics and engineering and other sciences, we become more and more able to escape from the bounds that nature has imposed. We first escaped from the Great Rift Valley, and then we escaped from being obliged to stay on the ground. And then we escape from our planet, and eventually we will escape from our solar system and from our galaxy, and so on. So, although in our parochial perspective we have only been able to affect a very tiny proportion of the physical universe, our potential effect extends over the whole physical universe and is very profound because the kind of things that happen in the physical world once we get at them are radically different and have no parallel in any effects that any other kind of physical object can have. So that's the physical effect.

Now I think the theoretical effect, the philosophical significance of humans, is even more profound, and this is best seen via constructor theory. That's my idea that the laws of physics are best expressed in terms of what physical transformations are possible or impossible, 'impossible' meaning forbidden by laws of nature and 'possible' meaning permitted by laws of nature. And so if you think of the set of all conceivable tasks, the ones that occur without the intervention of, let's say, living things at all are a very tiny subset of those that can happen with living things. For example, the chemical reactions that occur without living intervention in organic chemistry is a very tiny set of reactions compared with organic chemistry, and the same with construction of physical objects like nests and bodies of animals and so on. But then when it comes to humans, the type of physical transformation that can be effected once there is explanatory knowledge, which is the signature type of knowledge that only humans create—that we know of—then that is again larger by an astronomical factor. The number of transformations that is possible to achieve given explanatory knowledge is enormously larger and is equal to the set of all transformations that can be achieved by the laws of physics.

So in two different ways, humans are aligned with the physical universe as a physical object, and they are also aligned with the laws of the physical universe as laws.

6:31 **Micah Redding** There's something counterintuitive about that because you're saying that the number of things that can happen without life and without intelligence is, in a sense, smaller than the things that can happen with life or with intelligence.

6:51 **David Deutsch** Yes, tiny.

6:53 **Micah Redding** But I think most people would think that's backwards, that we are small beings, that we're not running a sun in our basement or something like that. There are things happening out in the universe that we're not doing specifically.

7:11 **David Deutsch** Yes, well, as I said, this is largely a matter of perspective. It's because in our experience in the past, we haven't had that much effect. We have only explored a tiny proportion of our potential in this regard. But as regards whether we have an artificial sun in our backyard, in a small way we soon will when we have controlled fusion. But the point is that there is a fundamental connection built into the laws of physics between understanding and affecting. So, in principle, anything that we understand, we can affect to the largest extent that is permitted by the laws of physics. The laws of physics don't permit us to make literally a star in our backyard because it wouldn't fit. And that's because of various laws of nature. But to make a star out of interstellar hydrogen is a task that's possible in principle. And if someone wants to claim that that's not feasible to human beings, that that transformation is not accessible to human intentions should we want to do it, then they'd have an uphill struggle because ultimately there's no way that they can make that case without claiming that there's a law of physics preventing it.

8:47 **Micah Redding** One way you've described this before I think is [via] our special relationship with the laws of physics. And that's a really interesting concept. I want to come back to that and the concept of knowledge and so forth. But let's step back. One thing that you've talked about in several different contexts is the idea that we actually don't live in a normal spot in the universe. That our place in the universe is a little bit unusual and that the usual places in the universe, the average places I suppose, are just completely dark and completely empty as we would normally perceive it. Is that an accurate statement?

9:40 **David Deutsch** Yes, that's an accurate statement about intergalactic space. And even interstellar space within the galaxy, while not quite as dark as that, is still a very cold and dark place compared with what we're used to. On the other hand, there is also the truth that even the best places on our planet for our evolution, which presumably the Great Rift Valley in Africa was one, is still extremely hostile to us. By evolutionary standards, it's favorable to us in that it allowed us to evolve and it didn't make us extinct before we did evolve. But that's a very low hurdle to jump by human standards. By human standards, people talk about Spaceship Earth when they try [to] make a metaphor for how well-suited the world is to us, but spaceships would fail every possible safety check if they were as dangerous as the Great Rift Valley where we evolved. It had vermin, it had disease, it had extremes of various kinds of weather which would kill us.Indeed, it did kill most of us before what we now regard as our natural lifespan.

11:16 **Micah Redding** And so you're saying that there is no environment that is particularly suitable for us?

11:30 **David Deutsch** Quite so. That's right. All the things that make it suitable for us, make it especially suitable for us, have been created by humans by way of first generating explanatory knowledge and then using that to affect the physical world and thereby change our environment to make it more suitable.

11:54 **Micah Redding** And so the argument that you make is that we've done this for ourselves in our history, on this planet, and that there's nothing in principle that stops us from doing it in an environment even as extreme as intergalactic space.

12:14 **David Deutsch** That's right. We take for granted the kind of changes that we've made on Earth that, for example, make it possible for me to be warm and have a pleasant conversation here in Oxford, England on an early spring day when the temperature outside would definitely kill me in a matter of hours if I didn't have technology to protect me. So we take that for granted. We take for granted that we can thrive and be comfortable in environments which our ancestors would have regarded as deadly and which would have killed them. And similarly, when we look to the future and think what it will be like to live on the Moon or Mars or [in] another solar system, we kind of view the environment there as intrinsically hostile compared with the Earth. But it isn't. It's only unfamiliar. The problem of living there is exactly the same as the problem of living on Earth. It's a matter of using the opportunities that knowledge and the laws of physics provide to automate the task of making living comfortable. And we've done that many times on this planet. We will do it many more times on other planets.

13:39 **Micah Redding** And in your book you talk about what we would be able to do even in the ultimate blackest place in the universe, that we would still have the resources we needed to, given the appropriate knowledge, create a habitable environment. And I guess that goes to what you're saying about our special relationship with the laws of physics. You're saying that there is nothing in this universe, there is no place in this universe that is in a sense beyond our reach because of that relationship.

14:23 **David Deutsch** Yes, almost nothing. I did say that I'm not sure about the interiors of quasars, whether we could live there. But yes, there's almost nothing beyond our reach. And what's more, looking at [it] the other way around, there is no other animal and no other physical object in the universe of which that is true.

14:43 **Micah Redding** And okay, so that comes down to a lot of what you talk about. What is it that gives us that unique relationship? What is special about that relationship? How is it that human beings are special in this way? When I think so many people in the scientific world would say that itself is just a kind of a biased, anthropocentric viewpoint.

15:16 **David Deutsch** Yes, well, one has to look at the facts. And what gives us this ability is—all these different possibilities we've been discussing come through a single ability, which is the ability to generate explanatory knowledge. That's kind of analogous, if you want to think of an analogy, to the fact that all the abilities of organisms, of living things on the Earth, including us in our capacity as just animals, come through a single capacity, namely the capacity to evolve knowledge in genes, in DNA, via variation and selection. So all the different abilities of living things, of flying and of generating chemicals inside their cells and harnessing nature in the way that living things do, all those come through that one ability to generate dumb knowledge, if you like, the kind of knowledge that is embodied in genes. And everything that is distinctively human, we have that as well, our species has that as well, but it doesn't begin to scratch the surface of what explanatory knowledge makes possible. That makes possible everything that is possible. So it's that single ability to generate explanatory knowledge that is responsible.

16:57 **Micah Redding** So you're saying that there is a kind of knowledge that exists in genes and that there is a higher order knowledge in some sense that exists uniquely in humans.

17:08 **David Deutsch** That's right.

17:10 **Micah Redding** This is something that I think maybe is the hardest bit to grasp. In your books, you talk a lot about knowledge as the most essential feature of our universe, the most essential thing that we could know about our universe. So how do you explain what knowledge is? Because you've talked about it as kind of a physical thing.

17:45 **David Deutsch** It's physical, yes, although it has properties that transcend any particular physical instantiation. So the kind of knowledge that is passing back and forth between you and me at the moment is being translated through lots of different physical objects which obey rather different laws. Part of it is encoded in moving electrical and magnetic fields as it passes from one continent to another, but it's also in sound waves and it's also in electrochemical neuronic signals. So information is a physical thing, but you have to understand that it obeys its own laws, just like electricity and magnetism do. All physical things are kind of unified.

Now, knowledge is a special kind of information. It's information which can be used to achieve physical transformations, and which is necessary to achieve most physical transformations that are possible. And it is, in a sense, the most important of physical phenomena, because to understand how stars work, you need to understand nuclear physics and gravity and hydrodynamics and so on, and to understand how fish work, you need to understand biochemistry and so on. But the overwhelming majority of things that are possible and that will happen eventually, to understand those things, to understand why those things happen, you have to understand knowledge. So an understanding of knowledge is the most needed understanding to understand what happens in the physical world.

19:58 **Micah Redding** Yeah, I think you used the example of looking at a star or something like that, and that the most important thing to know about that star is whether there are intelligent beings in the vicinity of it.

20:14 **David Deutsch** Yes, well, that is for most things you might want to ask [about], yes. So if you want to ask, "Will it go out soon?" You can answer, "Well, it won't go out soon unless there are intelligent beings there." If there are intelligent beings, you have to understand a lot more about it, because you have to understand a lot more about them, to ask, "Will they want to switch off their star or to interrupt its light traveling to us or not?" And that requires a completely different kind of knowledge from the knowledge that we currently expect to encompass everything that we know about stars.

20:54 **Micah Redding** Yeah, so life and intelligence and knowledge are really the most important factors in the universe in the development of how star systems evolve, of how even galaxies evolve. That is both a huge concept and, as I've already said, very counterintuitive to a lot of people. Can you talk about the concept of universal reach? In *The Beginning of Infinity*, you lay out this idea that there is a difference between things that have finite reach and things that have infinite reach. And I think that's a pretty unusual concept for most people, because it comes up in things like artificial intelligence. Will we be forever behind intelligences and all kinds of different things? You used an example of Roman numerals versus place notation as a way to describe that. Can you unpack that a little bit?

22:14 **David Deutsch** Yeah, so as you said, when we solve problems and create explanatory knowledge, usually the knowledge extends as far as the problem did, if we're lucky. I mean, we may not solve the problem, but usually the good outcome is that the knowledge extends as far as the problem, and then new problems will arise and we have to generate new knowledge. And this was the case with number systems. First they had tallies where the number of marks on a stick was equal to the number of sheep, and you could do certain operations about counting sheep—whether you have lost or gained someone, and so on. And then better systems were invented, like Roman numerals, where you didn't have to mark the rod a hundred times to indicate a hundred sheep. And eventually that evolved into a number system which is as good as you can get in that respect. It's as efficient a way of representing sheep and any other thing that numbers can be used for. It's as efficient as can be achieved using the laws of physics. So that's an example of universality in a number system.

I think you're really asking about the particular kind of universality that human thought has. Because there are many kinds of universality. One important one is the universality of computation, the fact that one computer can compute basically the same things as any other, limited only by its speed and memory capacity. But not in its repertoire of computations. But you're interested in human universality, is that right?

24:08 **Micah Redding** Yeah, let's talk about that, because that's something you've described in several different contexts, this idea of human universality, and applied it to all kinds of different questions. And I think what you're suggesting is not just that humans are this amazing thing, but in some sense that humans have reached a point beyond which there is in a sense no further ontological change, I guess.

24:50 **David Deutsch** There's no further change in capacity to understand and control the world, yes. Again we are at any particular time, like a computer, limited by the amount of hardware that we happen to have. But there's no limit to the amount we can make if we want to. So this human universality, well that's the universality in the distinctively human ability that we were talking about just a moment ago, about the ability to generate explanatory knowledge, explanatory theories. Now there are various ways, I argue in the book, against all the possible ways that one might think that there might be a fundamental limitation on that. And I argue that none of them make sense, all of them are equivalent to just arbitrarily limiting our knowledge with a supernatural edict.

So one of the arguments, for example, is [to] suppose there were a limit to how much we can understand. Martin Rees suggested that there might be aliens out there in space somewhere who are as far above us in their capacity to think—not in their technology of course, that could be arbitrarily far ahead of ours—but in their capacity to think as we are ahead of chimpanzees. And he seems to think that there's no reason to think there aren't such beings, whereas I think that the point is, because of the universality of computation, which is a lesser thing but which our brains certainly have that, because they are computers and can perform ordinary computations as well, because of that, and because computation is itself universal in that everything physical can be described by computations—we know that from the laws of physics—so all these things come together, because of those two forms of universality, anything that the aliens can think, can understand, could be represented as a computer program. And if it can be represented as a computer program, then it can be represented as a thought as well, a human thought, if necessary, by augmenting human brains with additional memory, which will no doubt be done for many other reasons very soon.

And in fact, to all intents and purposes, it has already been done, and has been done for millennia already, because when the first person invented writing, they were already augmenting the hardware of the human brain, its memory capacity, and also its computational capacity. And so when people throw up their hands at the idea of augmenting the human brain with computers and saying that such a thing wouldn't be human anymore, they are simply making the same mistake as thinking that you're not really human if you use a pencil and paper to do your calculations.

28:32 **Micah Redding** So you're saying that the fundamental limits that we might face, which are memory and computational capacity, are already things that we have worked to overcome and that's ultimately what technology, at least information technology, does for us. It extends our brains in that way, or the ability of our brains. And you're saying that, because we know that our brains are at least what a computer is, we then know that everything that's possible in the universe, every concept that's possible in the universe, can be modeled in our own brains.

29:25 **David Deutsch** That is correct.

29:26 **Micah Redding** I think is a lot to swallow for a lot of people. But how do we know something like this? How do we know that our brains are computers or that they have those kinds of functionality?

29:52 **David Deutsch** So as I said, that is a conclusion, not an axiom, not a premise. And it's a conclusion from our best knowledge. In that particular case, it's because of the universality of computation in the sense that we know that any physical process can be modeled with arbitrary accuracy by a computer. The reason we know that is because we know what the fundamental language in which the laws of physics are expressed is, namely the laws of quantum mechanics and relativity. They underlie all the other laws of nature that we know. Now, relativity doesn't really enter into it, but quantum mechanics imposes constraints on what can be computed and [it is] really responsible for the laws of computation. And I proved back in the 1980s that a universal quantum computer would be able to compute anything that any other physical object can compute, and that includes aliens. Of course, no one can prove that the laws of physics are quantum [mechanical]. [It's] just our best theory, as unrivaled at present. There's no rival theory. But then we're not in the business of proving things. Science proceeds by trying to explain things, and that is the best explanation that we have.

31:34 **Micah Redding** I want to jump into your analysis of history a little bit, because, like you've already said, you think that we emerged in an environment that was essentially hostile to us, that we were able to leave that environment due to gaining knowledge.

32:02 **David Deutsch** First to improve it, and then to leave it, and then to improve the environments we found, and so on.

32:08 **Micah Redding** But at the same time, you would say that for the vast majority of human history, we were making almost no progress until something changed.

32:24 **David Deutsch** Yes, well, we were making almost no progress in two senses. One is that it was sporadic, and it was sort of two steps forward, two steps back, quite a lot of the time. And the other way is that it was extremely slow. The rate at which knowledge was being generated was very slow compared with a human lifetime, and therefore a typical human would never see any increase in knowledge in their lifetime, for most of human history. And that only stopped being true a few hundred years ago. It's somewhat arbitrary, you know, with the scientific revolution, but I think the real key change was a bigger thing, namely the Enlightenment.

33:14 **Micah Redding** So what is it about the Enlightenment or the Scientific Revolution? What changed then that allowed all this stuff to start happening?

33:25 **David Deutsch** There are a number of different ways of expressing what changed. Karl Popper says that what is needed for progress is a tradition of criticism, which kind of sounds like a contradiction at first, because 'a tradition' means a way of keeping things the same, and 'criticism' means a way of changing things. But if you think of it as being a tradition of criticism, then that's a thing which, if you can achieve it, it stabilizes the changes, and then you can have the growth of knowledge. Now, I have said what specifically happened there, if you like, at the psychological level, or at the level of individual minds, to create, to implement a tradition of criticism, to make the conditions for a tradition of criticism to happen, is that people started looking for what I call good explanations, where by 'good' I mean that they are hard to change while still accounting for the things they purport to explain. People have always looked for explanations, and they've sometimes found them, and often they've been false, but what they haven't been able to do is systematically improve them. And once you look for good explanations, which is explanations that are hard to vary while still accounting for what they purport to account for, that means that they engage with the conditions of the problem that they're trying to solve. Once you're looking for that, then you can make progress that builds on previous progress.

35:21 **Micah Redding** This idea that explanations that are hard to vary, I think that I resonate with that. Maybe instinctively, I've had a kind of motto for myself that truth is the thing that you can't shake once you grasp it, or something. And so that seems to be similar to what you're saying, right? That this is a very particular, like it's an explanation that's very particular.

36:00 **David Deutsch** Yes.

36:01 **Micah Redding** So any kind of variation breaks it.

36:05 **David Deutsch** Yes, although there are other ways of stabilizing ideas, irrational ways. But yes, if your ideas are stabilized by their engagement with the problem, then that's a good explanation. By the way, even being a good explanation doesn't guarantee truth. We can be mistaken with good explanations, but if we continue seeking them, we will encounter problems with even our best explanations and can then improve on them.

36:38 **Micah Redding** So, you're saying that for whatever reason, something happened during the Enlightenment or during the Scientific Revolution which kicked off this culture of looking for good explanations and then a tradition of criticism around those good explanations.

37:01 **David Deutsch** Yes.

37:02 **Micah Redding** And that process itself is then the thing that means we can just keep going from here to infinity, basically.

37:17 **David Deutsch** That's right. I think we've passed the threshold. We've passed the jump to universality.

37:25 **Micah Redding** And that's a pretty big idea that some year in the last few hundred years was essentially the first year in the history of the universe, perhaps, in which this stuff became possible, but now it is.

37:44 **David Deutsch** That's not quite right, I think, because there's no guarantee that we will, just because we can. There's no guarantee that we will. There's no limit. There's also no limit to the size of the error that humans can make. And it's conceivable that we will end the Enlightenment, or even that we'll wipe ourselves out, like most species have done. We, unlike any other species, are capable of not wiping ourselves out. But that's only capable. And I speculate in my second book, *The Beginning of Infinity*, that many attempts to form a tradition of criticism occurred during history. I suggest ancient Athens and Renaissance Florence as two examples. But I think there may have been many more, and all of them were wiped out within a couple of generations. The thing we call the Enlightenment, or the thing that we call Western civilization, or whatever you call it, has definitely been the longest lived of those, and the most widespread. But that's, I think, all we can say. Although there is a fundamental feature of the laws of physics that says we can continue doing this forever, there's no law that says that we will.

39:25 **Micah Redding** So we've talked about history. Let's talk about the future. What is the way for us to ensure our future? How do we move into the future, deal with some of the big things that we're dealing with from climate change to artificial intelligence to whatever might come up? How do we best guarantee, I know there [are] no guarantees, but how do we best ensure that we keep moving, that we keep progressing?

40:07 **David Deutsch** Yes. So I think that the key thing is to carry on generating knowledge as fast as possible. Winston Churchill said, "If you're going through hell, keep going." We are bound to encounter problems. We are bound to encounter large problems, and we are bound to encounter unforeseeable problems. So that's all going to happen, for sure. The temptation for some people is to rein in research, progress of all kinds, technological, scientific, even moral knowledge, to rein it in for fear of the unintended consequences of making mistakes. But this is a terrible idea, because there is no way of avoiding mistakes. The universe doesn't provide that. The universe only provides a means of solving problems, not preventing them from happening, and especially unforeseen problems.

And in the case of the problems that are caused by mistakes, in general human mistakes, some we cause by nature, some will be caused by human mistakes, and especially those that are caused by human moral mistakes, i.e. malevolence. We have to take into account the fact that malevolent humans have creativity at their disposal as well. They, too, can generate explanatory knowledge, and so how do we make sure that the good guys continue to defeat the bad guys forever? Well, I think that the good guys, although we are the same, the good guys are the same as the bad guys in regard to capacity, inherent capacity. There is one advantage that good has over evil, which is that, as I said at a debate recently, the bad guys are wrong, and the enemies of civilization are wrong, and therefore they, unlike the good guys, have an inherent interest in preventing their ideas from changing. And that's why they are always worse at generating knowledge than the good guys. If we try to rein in our knowledge creation because of fear of the effects of technology or whatever, then we are destroying our only advantage, quite apart from it being useless anyway because problems are inevitable. But for those specific kinds of problems, it's exceptionally perverse to try to limit the growth of knowledge.

Therefore, we must get used to jarring changes. They will happen faster and faster. So long as we keep the tradition of criticism alive, we don't know what even that will look like in the distant future. At the moment, we have various traditions of criticism, like liberal democracy in politics, and capitalism in economics, and peer review in science—we have those institutions. The particular institutions are not going to survive forever, but the property of being a tradition of criticism has to survive or we're doomed.

44:14 **Micah Redding** You apply the same thinking to the democratic process, which is interesting. So explain that a little bit. How does that make sense? How is democracy like this process of knowledge generation that occurs in science or rationality?

44:36 **David Deutsch** This idea I more or less repeated directly from Karl Popper. His theory of politics is essentially the same as his theory of the growth of knowledge in science. The idea is that policies are theories, they are ideas.There is no guarantee that they'll be right or that they'll be true, and we must assume that they are not true and will cause problems which, unless they are solved, will end the political culture in question. So in politics, just like in science, we need a tradition of criticism, and it has the same paradoxical and yet possible feature that it does in science.

For most of human history, those two things never went together. Tradition and criticism never went together, but after the Enlightenment, they did. We have traditions which stabilize change. Again, [this] seems like a contradiction in terms, but our political institutions stabilize change. Now this has some implications for how actual institutions like voting systems and legislatures and so on are arranged, because people still haven't learned the lesson of Karl Popper in that they still want to judge political systems by the same criterion that they would judge a particular policy. That's obviously fatal for the possibility of change, improvement. So they want to make a system that is most likely to make the right choice, for example, to make the right policy, to elect the right leader, and so on. In Popper's scheme of things, you simply take for granted that no system can do that. You take for granted that there will always be bad policies, bad institutions, bad leaders, and the real problem is how to make a society that can change those things easily, as easily as possible, without the society itself being destroyed. And so he would judge a political institution, a voting system, always by a single criterion: Does it make it more or less easy to remove bad leaders and bad policies without violence or the threat of violence?

So again, we kind of take for granted that in our system, in the Western political systems—all of them—a politician, a leader, may have been in power for several years, has been used to everyone doing what he says, and is sure that rival leaders, with their bad policies, will ruin the country, or even the world. And then they lose an election, and despite still thinking that the rival policy is bad, they quietly leave office. Not only do they quietly leave office and shake hands with the new leader, they would fight and die to make sure that that leader, and not themselves, stays in power. That's an amazing thing that we have achieved. It's been achieved very, very rarely in history, and when it has been achieved, it's been soon destroyed. But it's very precious.

48:57 **Micah Redding** Yeah, that's really interesting. Yeah, that is true, where the handoff of power is in a sense more important than the power itself.

49:10 **David Deutsch** Yes, much more.

49:13 **Micah Redding** Let's talk about one thing that kind of connects to our earlier discussions. You wrote an article a while back in which you argued that AI is a philosophical problem. I don't know if I'm characterizing that correctly. Of course the world is full of artificial intelligence of various kinds, but you're talking about artificial general intelligence, and intelligence like our own. And you're suggesting that, contrary to what some people might think, we're not going to get there just by kind of cranking up the clock speed over time.

49:56 **David Deutsch** That's right, or even by making better algorithms of the kind that search engines have and game-playing programs have. The thing is, intelligence like ours, intelligence that's like ours, according to me, means that it has the property of being able to create new explanations. That's our distinctive feature. I think I called it [a] distinctively human feature, but even that's not general enough, because it might exist in aliens as well, and eventually it will exist in artificial intelligences. And my general term for the things that have this ability is 'people.' The only people we know of at the moment are humans. But one day, we will make artificial people who will have this ability as well, and they will be fully human in every sense automatically as well, just by having that ability, because it's universal. And the same with aliens. You either do or don't have that ability. We can talk about the dividing line if you'd like.

All the existing programs that are called AI are not capable of generating any new explanations. The easiest way to see that is by seeing that you can write a program to determine whether a particular other program has the ability in question. For example, if somebody purports to have a program that plays good chess, then you can have a criterion in another computer program for whether it plays good chess. But it's intrinsically impossible to have a criterion, to implement in a computer program, a criterion for whether somebody has generated a new scientific explanation. Because to have that criterion, you'd first have to have the explanation, and therefore it wouldn't be new.

52:16 **Micah Redding** To put this simplistically, what's the difficulty in just sitting down and writing the algorithm that's going to create that new knowledge or whatever?

52:34 **David Deutsch** For this same reason, every other computer program that we write, before we write the algorithm, we know what property the algorithm is going to have, like what the desired output is for a given input. For an explanation generator, we precisely can't do that because we don't know what the new explanation is before we have it. So the task is different. What is needed to achieve that, I don't know. We only know from the very nature of universality that there exists such a computer program, but we don't know how to write it.Unfortunately, because of the prevalence of wrong theories of the mind and of humans and of explanation and of theory of knowledge and so on, all existing projects to try to solve this problem are, in my view, doomed. It's not because they're not using the right philosophy, it's because they're using the wrong philosophy. We have to stop using the wrong philosophy and then use the right philosophy, which I don't know what it is. That's the difficulty.

54:06 **Micah Redding** So you're saying as well that this process that exists in our brains and that makes us persons is something we completely don't understand. We don't understand how it works. We know it's there, we know it can be implemented in a computer, but we don't know what it actually is.

54:31 **David Deutsch** Yes, we know some things about it, like you said. We know that it can be implemented in a computer. We know some of the laws of epistemology, but yes, we don't know how specifically human-type, person-type creativity works.

54:47 **Micah Redding** I'm curious, you mentioned that dividing line. Does that change your notion of personhood from maybe a more traditional notion…

55:06 **David Deutsch** Yes. So personhood is a property of a program. It's a property of a computer program or a brain program or whatever. Therefore, it slightly misses the point to ask things like, "Are chimpanzees human or are they almost human or do they have human-type, do they have person-type characteristics?" Because it's not the chimpanzee that would have them, just like it's not the human body—the human brain that has them. It's the program in the brain. So it's not inconceivable that a chimpanzee brain could in principle be programmed with a person program. It's just that the standard interface, as it were, for a chimpanzee brain doesn't make that easy. In fact, apparently it makes it very difficult. People have tried to instill human culture into apes by doing the same thing that we would do to humans, that we do to humans to give them human culture, and it hasn't worked. So it hasn't created anything capable of creating any explanation, any new explanation. That's not to say it's impossible, and if it's not possible by the conventional methods of bringing up children and so on, then it may be, and presumably is, possible by nanosurgery, by actually changing the computer program in a chimpanzee's brain. It has a smaller capacity than a human brain, but I don't think that would make the difference, because the capacity of the human brain has to last for 100 years or so, and there are people who reach the age of 100 without being intellectually impaired. [A] chimpanzee's brain is a reasonable fraction of ours, so maybe it would reach its physical limits sooner than ours, maybe it would reach them in ten years or in five years, in which case it would probably be immoral to insert this program into a chimpanzee's brain.

These speculations, though, don't really matter from the point of view of the question that people want to ask, namely, "Is it really true that chimpanzees and other animals, all other animals, are qualitatively different from us?" They are, in that the program that they currently have, or that it's currently feasible to put in them by reasonable means, definitely has no such ability.

58:05 **Micah Redding** And if I'm not extrapolating too far, some people, at different kinds of points in history, try to make claims about qualitative differences between the brains of people of different races, but you would say that that kind of difference, whatever there might be, is irrelevant, because the real question is essentially this software, this cultural software that we have.

58:38 **David Deutsch** Yes, it's strictly irrelevant. The universality trumps any such differences. It's rather like imagining that a gameplaying laptop couldn't run a word processor, or even more, it's like imagining that somebody could write a word processor on which you could only type right-wing articles and not left-wing articles. It's that much of a misconception.

59:12 **Micah Redding** There was one other thing that I wanted to ask you, and I don't know how deep of a well this would be, so we can pass this by if you like, but we've talked on this program about the Omega Point theory, and you have a kind of a complicated relationship with that, but you talked about that in your book, *The Fabric of Reality*.

59:57 **David Deutsch** Yes, but things have got more complicated since then.

59:59 **Micah Redding** Yeah, they have. But at the time, you made a kind of defense of it as a plausibility from the physical standpoint, I would hasten to say not the theological standpoint, but the physical standpoint. Can you just describe for me what your feeling about that is now?

1:00:23 **David Deutsch** The question that this is addressing, really, is this open-ended, this universality of the human condition, can it be extended to a literal infinity, so that literally [it] will never come to an end, and will always continue to improve, or is there some finite limit imposed by the nature of cosmology? And it used to be thought that cosmology must come to an end, essentially either because the universe will recollapse after the Big Bang, or because it will continue to expand until the matter is so sparse that computations can't be performed anymore. Tipler showed that in a particular cosmology that he called the Omega Point Cosmology, which, although it re-collapsed, an infinite amount of computation, literally [an] infinite amount, will be eventually performed before the end of the collapse, so that people will be objectively thinking faster and faster, without limit, and, subjectively, they can continue thinking forever. Now, that particular cosmology, although at the time it was quite plausible, is now believed by most cosmologists to be contrary to experiment.

The best theory now is that the universe will in fact expand forever, and that this is caused by a thing called dark energy, which we completely do not understand. So, we've gone from cosmology that was understood well enough to be able to say that the Omega Point Cosmology of Tipler is possible, and in my view was very plausible, to saying that our best guess is that that cosmology isn't possible, but also that we understand it much less now than we did before. It's a wonderful thing about science, that we can make progress by realizing that we don't understand something, as well as by realizing that we do. There have been suggestions that, in the dark energy cosmology, if that really is true, that we could achieve an infinite amount of computation anyway, by going slower and slower, and using the very dark energy that's expanding the universe to drive computation. So again, things [would] go slower and slower, but only at such a rate that the total amount is still infinite. So, that's a possibility, but I think anyone who pontificates about whether that's really true or not, isn't really up to date with the controversy. So, we don't know about cosmology at the moment.

I don't think it matters for practical purposes, because the amount of knowledge we're talking about is way, way past the limit of being comprehensible to us at present. Words fail me. I mean, it's so far beyond any conceivable planning horizon that it doesn't make any difference whether you think that the growth of knowledge will never end, or whether you think that it will only end in ten to the ten to the ten years.

1:04:19 **Micah Redding** Right. We can still plan our vacation and so forth. We don't have to...

1:04:25 **David Deutsch** Without expecting to run up against any limits.

1:04:29 **Micah Redding** Yeah. I think Tipler has more recently suggested that persons could engineer the collapse of the universe if we so desired.

1:04:43 **David Deutsch** Yes, that's also a possibility in some cosmologies, yes.

1:04:47 **Micah Redding** Okay. So, that might be possible. We just don't know. And at any rate, those kinds of even potential limits are so far removed from our potential as knowledge generating beings, we have infinite reach that extends across the entirety of existence as we know it, in a sense.

1:05:15 **David Deutsch** Yes, correct.

1:05:18 **Micah Redding** All right. Well, that's a great place to stop. I would love to go into your take on morality and so many other things here, but I think that's probably enough blown minds for today. But is there a good place where people can kind of keep up with your most recent publications or your most recent work? I'll put links to books in the show notes, but anything else you'd like to link to?

1:05:58 **David Deutsch** Oh, yeah, my website, I suppose, because everything else that's connected with me is linked from there. So, daviddeutch.org.uk I think it is. And yeah, there's also the constructor theory website, but my website links to everything, including my Twitter feed.

1:06:19 **Micah Redding** Okay, I'll put links to that and links to the books, which I definitely will recommend to everyone. But thanks, David, so much for having this conversation and hopefully we can reconnect sometime in the future.

1:06:35 **David Deutsch** Very nice and very interesting talking to you.

1:06:37 **Micah Redding** All right. We'll see you later.

JOHN HORGAN:

A PLAN TO DYE ONE'S WHISKERS GREEN

About the interviewer: John Horgan is a science journalist. He was a full-time staff writer at *Scientific American* from 1986 to 1997, when the magazine fired him due to a dispute over his first book, *The End of Science*. Eight years later all was forgiven, sort of, and Horgan wrote a couple of freelance articles for *Scientific American*, notably "The Forgotten Era of Brain Chips." From 2010-2022 he wrote hundreds of opinion pieces for the magazine's online edition.

Interviewer homepage: https://johnhorgan.org/

About the YouTube channel Nonzero: Conversations with a series of people who have nothing in common except that program host Robert Wright is curious about what they're thinking

EPISODE DETAILS

Date	April 5, 2018
Interviewer	John Horgan
Source	YouTube
Show	Science Saturday
Episode	Science Saturday: A Plan to Dye One's Whiskers Green | John Horgan & David Deutsch
Description	00:56 The father of quantum computation explains parallel universes 06:57 David to John: You're wrong, we're not nearing the end of science 23:25 WWET: What would Einstein think? 36:40 Abstractions are real things 46:48 Searching for the Charles Darwin of artificial intelligence 52:56 Unlimited progress and the future of civilization
Link	https://www.youtube.com/watch?v=y3Afrwe-Yyg

Ideas

- The prevailing way of trying to find fundamental theories in physics assumes that the future theory is going to still be based on things like particles, space, time, fields, and so on. Why should it be? Fundamental progress in the past has always involved new kinds of entity, new modes of explanation that weren't thought of before.
- The argument against free will from reductionism, the idea that all explanation must be in terms of microscopic things, is just a mistake. We have to find the best explanations that explain free will, rather than impose by dogma a criterion that explanations have to meet other than that they explain reality.
- We haven't solved the enormously important problem of how to transmit the knowledge that Western institutions consist of to political cultures that don't yet have it. And it seems that there's something about our existing political culture that is actually antagonistic to transmitting it outside its natural home. Unless the means of promoting the resolution of disagreements without violence can be propagated to basically the whole world, we're going to be in increasing danger from things like weapons of mass destruction in the hands of terrorists.

Topics

abstractions • artificial general intelligence • artificial general intelligence running on quantum hardware • Bohr • Darwin's theory of evolution • DeWitt • Einstein • emergent properties • environmentalism • Everett • free will • Galileo • global warming • good explanations • heliocentrism • information as a substrate-independent phenomenon • multiverse • optimism • Penrose • Popper • progress • quantum computation • rapid progress as means of good guys beating bad guys • realism • reductionism • string theory • the end of science • theory of everything

Episode title and description reproduced verbatim from the original source.

Transcript

0:01 **John Horgan** Hello, David Deutsch, are you on the other end there?

0:04 **David Deutsch** Yes, hello.

0:07 **John Horgan** It's great to have you here on Bloggingheads, David. Let me just introduce myself, and then I'll let you do the same. I am John Horgan. I'm a science journalist and occasional science correspondent for Bloggingheads TV, and I have a really special guest with me today, the British physicist David Deutsch, who is now in, you're in Oxford right now?

0:30 **David Deutsch** That's right.

0:33 **John Horgan** So David, could you just give us a little background on yourself and then we'll start talking about your wonderful new book.

0:41 **David Deutsch** Certainly. So my name is David Deutsch, and I'm a physicist at Oxford University and I'm the author of two books, *The Fabric of Reality* and recently *The Beginning of Infinity*.

0:56 **John Horgan** David, you are often described as a pioneer of, or even the father of, quantum computation, and that's been a theme in both of your books. Can you just describe a little bit about how you got into that field, which I think was quite a while ago, a couple of decades ago?

1:20 **David Deutsch** Yes, I actually began thinking about quantum computers in the 1970s, although I didn't call them that then because I didn't think of them as being anything to do with the foundations of computation. The context there was the so-called parallel universes or many-universes interpretation of quantum mechanics and I had realized that the consensus view that both the proponents and the opponents of this rather controversial interpretation had been taking, namely that it is just a matter of interpretation and that there are no possible experimental tests of it, was actually false, and that this idea that it can't be tested was simply due to some poor thinking about what would happen when an observer did a measurement on another observer and so on. And so I was trying to clarify this issue of what an observer is in quantum mechanics. So I thought, "Well, the simplest way to clarify that is to imagine an artificial observer, what would be called an AI or an AGI, an artificial general intelligence, but running on hardware that obeys quantum mechanics."

Of course all hardware obeys quantum mechanics, but I was thinking of hardware that obeys it in a way that can be tested in the laboratory. And so I imagined an [AGI] program running on this quantum hardware, and then I added a few extra elementary operations to this computer. I had to add a couple of extra operations to this computer because that's what made the difference between doing this experiment with a computer that obeys quantum mechanics and an ordinary computer, the kind of computer that we're familiar with. And given those extra operations, it was possible to perform an experiment whose outcome would be one way if there was only one universe—if something like the Copenhagen interpretation, the wavefunction collapse interpretation, or any single universe interpretation of quantum mechanics were true—and would go another way if the many-universes interpretation was true. And so as a sort of side effect of this, I realized that quantum mechanical computers would be inherently more powerful,could perform more qualitatively different computations, than classical ones.

4:42 **John Horgan** Maybe you should just back up and we can't assume that our listeners are completely familiar with all the different interpretations of quantum mechanics. Just remind us of what the many-worlds or many-universes interpretation of quantum mechanics is and how it differs from, say, the Copenhagen interpretation if that's kind of the more mainstream view of what quantum mechanics means.

5:14 **David Deutsch** To explain what quantum computers are, I have to explain what quantum mechanics is from my point of view. And I adhere to what's called the parallel universes or many-universes interpretation proposed by Hugh Everett in 1957. It says that the universe we see around us is just a tiny facet of the whole of physical reality, and so if we want to retain the same word for universe, we have to invent another word for the whole thing, and I favor the word 'multiverse' for the whole of reality. This, in my view and in the view of its other proponents, is an incontrovertible implication of quantum theory, which is our most fundamental theory in physics. But I always have to warn the viewer immediately that this view is shared by perhaps fewer than ten percent of theoretical physicists. At any rate, that's what I take quantum theory to say. And quantum computers are computers that harness quantum theory to perform a different mode of computation, something that cannot be performed by existing classical computers at all.

6:45 **John Horgan** I see. Okay. I want to come back to multiverse theories and the multiple universe interpretation, but let's talk about your book now. I'm holding up the galleys that I got from The Wall Street Journal when I reviewed your book, *The Beginning of Infinity*. And I should say that the journal I think asked me to review your book because it is almost the antithesis of a book that I wrote in 1996, *The End of Science*. And in fact, at the very end of your book, you mentioned my book and [rejected] its claim that science might be approaching its limits very forcefully. I think the journal expected me to lay waste to your book. They thought that that would be entertaining. And I fully expected to do that when I started reading you. But I ended up really loving your book. I think it wasn't what I expected at all. It's a really grand vision of human possibility, and it got me questioning my own pessimism about the future of science and even technological progress. So could you give us a nutshell version of the book's theme and also give us some sense of where the ideas came from?

8:24 **David Deutsch** Yes. The basic theme of the book is that all human progress in the past has been fundamentally caused by a single kind of activity which I call the quest for good explanations. Explanations [are] accounts of some kind of reality, how it works, and why. And pursuing this theme of what an explanation is, why the quest for good explanations can work and so on, makes contact with other bits of science and philosophy which together imply that this process need never come to an end. That is, we could bring it to an end if we destroy ourselves or decide not to or whatever, but there are no inherent limits to the growth of knowledge and therefore to progress.

By the way, you mentioned your review of my book. I thought it was an exceptionally nice, generous review. But it's funny you should mention your feelings on reading it because they were exactly mirrored in my feeling on reading your book. I was expecting to hate everything in it, but instead I merely disagreed with the conclusion. Correct me if I'm wrong about this, but it seemed to me that in every argument in your book there is a sort of reluctance, there is a wish that it were otherwise, and that your arguments about these limits, you are forced to them because you think that that is the logic of the situation, but you would rather that it were otherwise. And that's what I got from your book.

10:24 **John Horgan** Well of course. I mean, I became a science writer because I see science, as I think you do, as by far our most powerful way of understanding ourselves and understanding all of reality. And I actually got into science journalism in the early ’80s when there was talk of a theory of everything. Stephen Hawking had predicted the end of physics, [that] there would be this great revelation at the end of our quest to understand everything. And I was enormously disappointed when, after a period of time, I started suspecting that maybe science was already bumping into walls and we wouldn’t get these great revelations in the future. So you’re absolutely right.

11:16 **David Deutsch** I think these walls are of our own making. They’re not inherent in the subject of physics. The book covers all subjects, by the way, not only science. I think this thing about the quest for good explanations has been responsible for all progress such as moral progress, political progress, artistic progress, every kind. But science is my field, and physics in particular, and it is true that progress in fundamental physics reached an all-time high in the early twentieth century, and although it’s by no means gone to zero now, it is lower now than it has been in the past. And this has caused some people to think that either we’re reaching the end of all knowledge so that we’ll understand everything, or we’re reaching the end of the capacity of science to create knowledge. And either way we’re heading for a brick wall. I think this apparent brick wall, as always in the history of knowledge, was not caused by anything in the subject. It was caused by what people have chosen to do.

You mentioned the theory of everything. To me, [this] proposed or hoped for theory of all elementary particles, space, time, and gravity should not be called the theory of everything. That is a very tiny facet of physics from my perspective, let alone everything. It’s just the theory of how objects behave. But beneath that, all such theories are formulated within a certain profoundly significant language and conceptual framework, namely quantum theory. And the theory of everything just assumed that quantum theory would survive, would be exactly the same theory after we have discovered this great unification. And essentially that the theory of gravity also would, and that unifying them would simply be finding a way of writing either of them in the language of the other.

That has been the technique of elementary particle physics for the past several decades. They're trying to formulate a classical theory, not counting quantum mechanics, as if it was all in one universe, as if there weren't interference phenomena and tunneling and all those things…just a classical theory. And then they apply a process that physicists call quantization, which is a way of transforming a classical theory into a quantum theory. So you turn the handle, it's just a mechanical process. This worked for electrodynamics. That was the great achievement of Richard Feynman and Julian Schwinger and others. It really hasn't worked since, and I don't think there's any reason to believe that this process will ever work again. It was just a stroke of luck that quantum electrodynamics can be obtained from classical electrodynamics by a mechanical process of quantization, plus a whole load of cleverness.

15:16 **John Horgan** Does this mean that you're skeptical of string theory and even loop space theory and some of the main contenders for a theory that could unify relativity and quantum mechanics?

15:33 **David Deutsch** Yes, unfortunately. Although I wouldn't say these things aren't worth doing because, if nothing else, we learn from them even when they fail. But it seems to me that because progress comes from good explanations, it has to come from problems, because an explanation is an explanation of something like how a thing can possibly be. And that means that the prevailing way of trying to find fundamental theories in physics is unlikely to succeed, because it is looking for mathematical models and then trying to understand what that could possibly mean if it were a theory in physics. Even if you found the right theory that way, I think the chances are fairly low that you'd recognize it, because how do you know which of those mathematical objects correspond to which objects in nature? We're assuming that the future theory is going to still be based on things like particles, space, time, fields, and so on. Why should it be? Fundamental progress in the past has always involved new kinds of entity, new modes of explanation that weren't thought of before. So yes, I'm skeptical that these approaches, any of these approaches, can work, and I think that's why there hasn't been this fundamental progress.

17:07 **John Horgan** David, let me raise an objection to your optimistic vision of the future of science, which is actually based on my reading of what quantum mechanics has done in physics. You have a passage in your book where you recall, I think it's Niels Bohr, saying that anybody who thinks he understands quantum mechanics obviously doesn't, and you reject that as a kind of know-nothingness, which is surprising for someone like Bohr. But it seems to me that if you look at, in a sociological sense, all the different competing attempts to understand what the hell quantum mechanics means, you'd have to grant that even for the experts, the theory is quite confusing. On the other hand, it's powerful. It does anything that you could want from a theory in terms of being able to predict experiments and lead to all sorts of amazing applications and so forth. And so it seems to me that you're getting a split between science as giving us power over nature and science as a mode of understanding.

The understanding, and especially then if you look at the rest of physics as well, has become, for the average person, extremely difficult to understand, very esoteric. I see science as really, beginning in the twentieth century, becoming more and more distant from the comprehension of the average person. And I just wonder where your optimism comes from that somehow in the future, I don't know, as a result of new ideas in physics, new ideas about how unification should take place or whatever, why we should get the comprehension that seems to be retreating now.

19:11 **David Deutsch** In my view, this split that you talk about between quantum theory as a powerful technique for building things and making predictions, and quantum theory as a way of understanding nature, is not a feature of quantum theory. It is a feature of the sociology of science during the twentieth century. The split was introduced as a matter of philosophical dogma in order to protect from criticism the bad explanations that the founders of the theory and subsequent physicists have favoured for quantum theory. By the way, from the many universes point of view, they are all equivalent to saying, "Well, at some point when we're not looking, all the universes but one suddenly disappear, and we can't notice this because we're not looking." And that sort of thing. The response to careful, considered criticism of this view is to say, "Oh, well, you don't understand quantum mechanics."

As Hugh Everett, the founder of the many-universes theory, pointed out in a famous letter, we have been here before. A radical change in worldview was occasioned by the discovery of the heliocentric theory, that the Sun and not the Earth is the centre of what we now call the solar system. Galileo championed this theory. And in a famous conflict between him and the Inquisition, they tried to force him to renounce the theory. But if you look in more detail, what they were asking him to renounce was not the power of the theory, not its ability to predict. They were quite willing to allow him to espouse and to teach and so on. What they wanted him to reject was the claim that this described reality, that this described the solar system. And the new vision of the solar system that was entailed by the heliocentric theory was a jarring change from what had gone before, because, for example, it meant that the Earth beneath our feet, which is the paradigm of something fixed in common sense, is actually moving very fast. It's moving at a thousand miles an hour around the Earth's axis, and also moving around the Sun. We can't feel this because the laws of physics are constructed in precisely such a way as to cancel out any feeling that we might have about this motion. People, at first sight, [thought] this is a ridiculous idea, because it's like what Lewis Carroll said, "I was thinking of a plan to dye one's whiskers green and always use so large a fan that they could not be seen." It was accepting one thing just in order to make it invisible and then explain something else. And it's only when you look very carefully at what the theory says that you see how much and how enormously better an explanation it is of the observed motion of the planets and so on. And then it allowed further unifications by Newton of celestial mechanics and terrestrial mechanics and so on. So this split at that time was an invention of the Inquisition.

The split in quantum mechanics, regrettably, was an invention of its very founders. They didn't want to take the theory seriously as a description of reality.

23:24 **John Horgan** Do you think that Einstein, if he was alive today or had lived long enough to see Hugh Everett's theory, would have embraced it, or would [it] have made him even more frustrated with quantum mechanics and convinced that it had to be incomplete or wrong in some way?

23:43 **David Deutsch** It's hard to predict what an actual person would have said. But if we look at what Einstein wrote about quantum theory and his famous criticism in his great debate with Niels Bohr about quantum theory, all his criticisms are straightforwardly met by the many-universes interpretation. So he only missed it by two years and it's very, very frustrating. Bryce DeWitt, in his famous article introducing the many-universes interpretation, in a footnote he says that Einstein would surely have liked this. I think he would as well, because what was driving Einstein both in theory of relativity and in his critique of quantum theory as it was in his day, was realism. He understood science and physics as being the study of what reality is like. And these equivocations that the quantum theory appeared to bring, namely, "Well, what do you really mean by 'real' and we can't really say what's real, we can only say what we observe about the reality," and so on, in which case science becomes the study of us. It becomes the study of our perceptions, and everything else is just a sort of fiction. That he rejected rightly.

Having, by the way, believed stuff like that in his youth and [rejecting] it in order to make progress with relativity, he then applied that idea of realism to the whole of science and insisted on that and rejected the quantum theory of the time as not being realistic, whereas the Everett theory is entirely realistic. In fact, you can define many-universe quantum theory as just the statement that the equations of quantum mechanics describe reality. That's all it is.

25:54 **John Horgan** Let me bring up another possible objection that Einstein might have had. Einstein has this wonderful phrase, I have no idea when he said it, but that the goal of physics is to determine whether God had any choice in making the universe. It's a way of getting at the question of, "Even after we figure out the laws of the universe and its history and so forth, why this universe? Why do we find ourselves living in a universe that allowed our existence?" and so forth. It seems to me that, and there was a hope, Steven Weinberg has also talked about this, there would be a theory at some point that would be kind of logically inevitable or necessary and if you tried to tweak it, it would fall apart and it would make this universe that we live in also necessary or inevitable in some sense. What has happened over the last couple of decades is that things have gone completely in the opposite direction, and now you have theories that predict basically an infinite number of other universes. So you're sort of back to the arbitrariness and the problem, as far as I can tell, has gotten even worse. So I wonder how you respond to that issue.

27:22 **David Deutsch** So first of all, if one interprets Einstein's view, that quote of Einstein's as saying that we need an ultimate explanation, then I think that that is a chimera. I think that that will never be found and can never be found, and if such a thing could be found it would be a catastrophe. It would be the end of progress, and progress in science is intimately connected with all the other kinds of progress, so it would also lead to the end of progress in the other ways that we like, such as morality, politics, and so on.

But I think there is nevertheless a truth in it, which is the truth about good explanations. What we want from a good explanation, in the way I describe in the book, is that it be hard to vary. That is, if you displace one note, as Schaffer said in the play Amadeus, then there's diminishment. And if you displace a phrase, then the whole structure falls apart. So in that sense, a good theory must have a certain inevitability about it, with hindsight of course. With hindsight, you see that it couldn't have been any other way. But what saves us from the evil implications of an ultimate explanation is that good explanations solve the problems that they address, but they always raise new and better problems.

So the problems that we have today, for example in cosmology about what the dark energy is that's making the universe expand at an accelerating rate, one of the most startling discoveries of science in recent times, that discovery depended on the previous discovery of the general theory of relativity and cosmological models in that theory and so on. It's only in the light of those theories that we can even know that the expansion of the universe is accelerating and know that that's amazing. So in solving the problems that Einstein solved, namely how things like the motion of light and the existence of gravity could be reconciled in his general theory of relativity, that opened up problems that were simply inconceivable before. One couldn't have expressed them even in the language of physics or in the language of common sense. They were problems that opened up because of the solution of previous problems. And that's the solution to the conundrum, "How we can get our hard-to-vary good explanations without grinding to a halt as a result?" It's because good explanations open up new problems.

30:34 **John Horgan** Let me bring up another figure who's very prominent in your work, the philosopher Karl Popper, who I was fortunate enough to interview a few years before he died. Obviously, he's been a very big influence on you. And I just wonder how Popper would have reacted or did react. I don't know if he ever wrote on multiverse theories, but I would suspect that Popper would have been a skeptic of multiverse theories because he was so insistent on testability. And it seems to me that multiverse theories are, at the very least, extremely difficult to test and that any kind of evidence you would have of their existence would be circumstantial at best.

31:33 **David Deutsch** Two issues there, one about the testability of the many-universe interpretation and the other one about Popper. I was also privileged to meet Popper on one occasion when I was a student and was lucky enough to be invited along to a meeting between Popper and my mentor [and] physics boss, Bryce DeWitt. And basically at that meeting, DeWitt told Popper that he had misunderstood what the fundamental problem is in quantum theory. Popper thought it was to give a meaning to probability statements, and he'd kind of missed the deeper problem of things like entanglement and interference and the measurement problem. Popper said that he had realized that he had an inadequate understanding and had held up publication of one of his books in order to try to improve it. I've looked at his subsequent books, and they all contain the same misunderstanding of quantum theory, unfortunately. He does occasionally mention the many-universe interpretation but only to dismiss it for kind of non-philosophical reasons, just to say, "Well, you know, we can't have that, and therefore I'm going to concentrate on this other thing." So Popper, unfortunately, like Einstein, died too soon, but not too soon chronologically. He just died too soon for the right understanding to have reached him. I suppose that is not a coincidence, because ninety percent of the physicists whom he might have asked about the foundations of quantum theory would have given him nonsensical answers.

33:26 **David Deutsch** As for the testability of the many-universe interpretation, as I said, there are, in principle, tests that would test it against the rival theory that there is only one universe, but really that is grossly understating the scientific status of the theory. The reason [is] that one doesn't normally test an interpretation. Normally in science, one says that "Yes, indeed, the equations of the theory do describe reality." And it's really only in the case of quantum theory within physics that somebody has said, "How can we test the interpretation by itself?" Namely, "How can we test the statement that these equations, though we're not disputing that they correctly predict experiments, in fact represent reality rather than just what we see in reality?"

As I said, we've been here before in physics, namely at the time of Galileo and the Inquisition. But in the present day there is a very close analogue of this, and that is the Creationists who say that no one's ever seen a dinosaur, just like no one's ever seen parallel universes. All we have is the circumstantial evidence of fossils and the interpretation of fossils as being the remains of dinosaurs. Similarly, no one's ever seen parallel universes, but what we have seen is interference phenomena and the interpretation of interference phenomena as being due to the interaction of different universes, and there's no other explanation. So if you want to say the other universes don't exist, you have to do it by fiat, rather like the people who say that the world was created 6,000 years ago with fossils. So similarly, the conventional interpretations of quantum mechanics say that at the moment of a measurement, all the universes except one disappear, don't exist. And no one can contradict this because no one can see them.

35:54 **John Horgan** David, I'll just tell you, I think that's a stretch to compare doubters of parallel universes to—and I count myself one, not a doubter but more an agnostic and someone who thinks that it's kind of a moot issue because we'll never have good evidence— Creationists who aren't sure that dinosaurs really exist. But I want to get on to another really big topic that you raise in your book.

36:24 **David Deutsch** Okay. By the way, I was only saying that the logic is the same. The psychological motivation is not the same, but the logic is the same because of the existence of good explanations in both cases. But okay, continue.

36:38 **John Horgan** All right. A really wonderful theme that emerged at a number of places through your book was, and I'll put it in my own words and you can tell me if I've gotten it wrong, was a kind of critique of simple reductionism or materialism, which obviously is the prevailing philosophy of physics, that good answers will come from going to smaller and smaller scales and also focusing on things, on objects, on particles, and so forth. And you seem to be saying that that is a much too restrictive form of explanation, and that we have to recognize that what we might even call immaterial phenomena that aren't reducible to specific physical objects or processes can have a profound impact and have had a profound impact on reality, particularly our human reality, human history, the world of politics and culture and so forth. And it seems to me it's almost a rebuke of physics as the best mode of understanding the world. And you're emphasizing how important mind is and ideas are and thoughts and so forth. So talk about that a little bit.

38:18 **David Deutsch** Yes. Of course, I am a physicist and I'm profoundly opposed to any idea of non-physical explanations that contradict physics. So that's a no-no and really doesn't make sense. However, there are ways in which both emergent properties such as minds and life and so on have an effect and, as you said, also abstractions.

The fact that the theory of good explanations led to the idea that abstractions are real things was slightly surprising to me. I wasn't expecting the link, at least wasn't expecting it to be so strong as it is. But if you think about how to explain events, physical events like a footprint on the Moon, how do you explain how that happened? Well, it happened because of human ideas, of science. And you could say in this reductionist sense that, as you rightly say is the prevailing mode of explanation and the prevailing idea is to look down on other modes of explanation, those ideas are nothing more than configurations of atoms. So some physicists, some rocket scientists, put their brain into certain configurations of atoms, and those atoms then acted on other atoms which acted on other atoms which then ended up making a footprint on the Moon.

What that misses is the explanation of why certain configurations of atoms put footprints on the Moon, while the overwhelming majority of configurations that human brains have been put into in history do not have that effect. And it's because there's a certain type of information. This information can't, in my view, be reduced to statements about atoms because if you think about what that information does, it is in brains but the same information then gets transferred into, let's say, sound waves in air and then it gets transferred into ink on paper and then it gets transferred into magnetic domains inside a computer which then control a machine that instantiates those ideas in bits of steel and silicon and so on and so on. There's an immense chain of instantiations of the same information, and it's only special kinds of information that have this property that they are preserved and instantiated in successive physical modes. So what is being transmitted, what is having the causal effect, is not the atoms but the fact that the atoms instantiate certain kinds of information and not other kinds. Therefore, it is the information that is having the causal effect. If a particular instantiation of that information were damaged, then processes would come along to fix it, whether or not they could fix the physical instantiation. For example, if the computer goes wrong then we don't use the corrupted information. We go back and rescue the information from a different computer and we throw away the atoms that at one point instantiated it. So the information causes itself to remain in existence.

I think there's no way out of that mode of explanation, and if explanation is going to be the fundamental thing about our criterion, for example, about what is or isn't real, then we have to say that information and this particular kind which we call knowledge is real and really does cause things.

42:46 **John Horgan** It seems to me you bring up the word "choice" at a number of places in your book, and you emphasize the power of human choice. It seems to me that what you are really doing is defending the concept of free will. Maybe you can tell me if I'm wrong here, and, as some of the listeners out there know, I am a free will fanatic. I'm very upset that some prominent scientists recently have said that free will probably doesn't exist. It's an illusion. Stephen Hawking has said as much. Einstein, in a couple of his quotes, suggests that free will probably doesn't exist. So you believe in free will, I take it.

43:45 **David Deutsch** I certainly do. And I think that the argument against free will from reductionism is just a mistake. It's a fundamental mistake. It's the idea that all explanation must be in terms of microscopic things. There's no philosophical argument in favor of that that I'm aware of. It's just an assumption. It has historical roots in how science centuries ago escaped from the clutches of the supernatural. As I said earlier, certainly I'm opposed to any kind of modes of explanation in terms of immaterial things, in terms of abstractions, that contradict physics. But the idea that all such explanations by their very nature contradict physics is simply false. I just gave an explanation of footprints on the Moon in terms of the ability of certain types of information to preserve themselves in existence and so on, whereas other kinds don't, that I defy anyone to reproduce in terms of atoms. And I also defy anyone to show how that contradicts an explanation in terms of atoms.

We have to accept the physical world as we find it. We have to find the best explanations that explain it, rather than impose by dogma a criterion that explanations have to meet other than that they explain reality. So I think this fashionable reductionism is just a mistake. I'm sure that free will exists.

However, I think free will is one of a constellation of emergent, abstract—we're not sure exactly what proportion of free will is abstract or emergent—properties that are not yet understood. Things like consciousness, creativity, choice, free will, and so on. We do have good explanations about them at the emergent level, but we don't understand them well enough to make artificial ones. And I say in the book that my criterion for judging any theory of consciousness, free will, and so on, is "Can you program it?" And if you can't program it, then I cannot take seriously your theory of it. Now, I don't have a theory of it. I only have a theory that it exists. If someone says that it doesn't exist because we can explain everything without invoking it, I want to see those explanations.

46:47 **John Horgan** Roger Penrose, I assume that you know him, has proposed a solution to the mind-body problem involving quantum mechanics working in some way. To me, the mind-body problem and free will, which is obviously a big part of it, is the biggest unsolved problem in science. And people are just grasping at straws right now. So I just wonder if you see any kind of bluesky ideas that might provide a kind of framework for understanding it. Maybe also information theory, which some people have also tried to bring into physics.

47:34 **David Deutsch** Well, as you said, Roger Penrose is looking for a new theory to replace quantum theory, which would not only be a better theory in physics than quantum theory is, but would also solve problems like the existence of free will and creativity and so on. I'm pretty skeptical, for the same reason that I'm skeptical of the mathematical approaches that are currently fashionable in fundamental physics. I think that one solves problems in physics by addressing problems that are in physics, rather than [by] hoping that they have certain attributes, finding a theory with those attributes, and then hoping that it applies to physics. I could be wrong, but at present there are no such theories.

I think existing approaches to [AGI], artificial general intelligence, are all philosophically flawed, and I think that's why they haven't succeeded for decades. A philosophical advance is needed, and they are trying to get the answer without making any philosophical advance, and that leads them essentially to behaviorist models, and behaviorist models are non-explanatory models. They are models that just try to relate output to input without explaining why the output comes from the input and so on, and I think that that approach can't succeed, and it's the reason that this quest for [AGI] has not gotten anywhere during the last decades. What we need is first philosophical progress in understanding how creativity—I think that's the key thing that relates all these unsolved problems about free will, consciousness, and so on— is implemented.

We know a few things. It has to be, in the broadest sense, an evolutionary process. It has to work by variation and selection, or as Popper calls it in the case of science, conjecture and refutation, or conjecture and criticism. But we need to know the details, and the devil will be in the details. My guess is that once we understand what it is, we will be able to program it.

I think there's an analogy here with Darwin's theory of evolution. Darwin's great contribution, in my view, is not his scientific theory of evolution, it is the philosophical progress that he made in inventing a new mode of explanation. Not just a new explanation, but a new mode. Previously, everyone who had addressed the question, "Why are animals the way they are? Why are there adaptations there?" by supernatural explanations and scientific explanations, but all of them took for granted that what you had to do is find a reason why there are elephants. Why [do] elephants have long trunks? That kind of thing. And Darwin realized that that is a bad way of approaching the problem. To understand why elephants have long trunks, you must not ask why they have them as your initial question, you must ask what kind of process could give rise to trunks. And then, that they have purposes—some biological features have purposes, some have anti-purposes, like the peacock's tail—that all comes out in the sophisticated elaboration of the basic theory of how it could possibly happen, by variation and selection. By random variation, that is, undirected variation, and then directed selection.

Free will, consciousness, and so on, definitely involves that as well, but it involves something else that we don't yet understand, which it will take a new Darwin to realize. It took many decades between Darwin and DNA, I think it will be much faster, in between the person who discovers the correct philosophy of [AGI], and the programming of [AGI] will be a matter of months, not decades.

52:48 **John Horgan** I hope I live long enough to see that. That would be very exciting. In the limited time we have left, I want to make sure that we touch on some of the political themes that you raise in your book. What I enjoyed about the book was that it was so broad, and you had these very powerful ideas, especially about accepting our fallibility as kind of a mode to constant self-improvement, and applying that in all these different fields—science, culture, and politics.

So when it comes to politics, I wanted to ask you whether or not you think that, you know, Francis Fukuyama had this, had a book called *The End of History*, where he's saying that in a very broad sense, democracy plus free market capitalism represents the best we can do as far as finding a way of organizing ourselves. And I wonder, although of course there's a lot of tweaking we can do, I just wonder if you agree with that, or if you think that there could also be infinite progress in the realm of politics?

54:07 **David Deutsch** The same arguments that I use in the book for everything else apply automatically to politics and imply that infinite improvement—unlimited improvement is a better word—is possible there, too. Liberal democracy plus free market capitalism is our best existing knowledge of this. And so I would guess that Fukuyama, despite recent hiccups in his predictions, is right that the ideas that had been the main rivals to those ideas during, let's say, the early twentieth century, such as totalitarianism, communism, and so on, that those are going into the dustbin of history. I think that is very different from saying that our best guess as to how to create new political knowledge is going to be just our current institutions. I'm sure that unlimited improvement is possible there, too.

For one thing, we haven't solved the enormously important problem of how to transmit this knowledge to political cultures that don't yet have it. And it seems that there's something about our existing political culture that is actually antagonistic to transmitting it outside its natural home. So that will be a major improvement, because as Martin Rees said in his recent book, in which he predicts that there's only a fifty percent chance of civilization surviving the next century, progress in other areas, especially technological areas, mean[s] that smaller and smaller numbers of people are going to be able to do larger and larger amounts of damage. And so unless the means of promoting the resolution of disagreements without violence can be propagated to basically the whole world, we're going to be in increasing danger from things like weapons of mass destruction in the hands of terrorists. By the way, I think this is not just a problem with improving our political system. It's a general problem to do with technology and everything else.

I should say that our civilization, the civilization of the West, of liberal democracy and capitalism and so on, is within itself by far the most peaceful as well as the most rapidly progressing civilization that's ever existed. But I think that apart from having to improve it further in order to allow it to survive, there's another thing we have to do. And this is a big theme of my book as well. We have to continue to make rapid progress. And it's not just for its own sake, but in this political context, it's because rapid progress is the basic means by which the good guys can defend themselves against the bad guys. I've said that technology makes a smaller and smaller number of people able to cause larger and larger effects. Well, that has to be offset by the larger number of people, the good guys, making at least as much progress as that in order to be able to cause even larger effects in self-defense. So it's rapid progress that is our major means of self-defense against the instabilities caused by small numbers of bad people.

58:25 **John Horgan** Okay, wait a minute. I've got to stop you there. That sounds to me like more arms races in the future. We've already been down that path and produced nuclear arsenals capable of destroying all life on Earth many times over.

58:43 **David Deutsch** That's not the implication of what I was saying. That's to interpret it in terms of the technology of the past. It's a sort of reductionist interpretation, if I may say so. Protecting ourselves against nuclear attack during the Cold War was done, and rather imperfectly done, by developing ways of nuclear attack ourselves. But protecting against, let's say, biological warfare attacks, basically what we need in the case of biological warfare is antidotes, not weapons.This is going to be increasingly so as the complexity and knowledge in society become the thing that we need to protect. For example, once we are able to download our minds from our brains into computers and so on, then physical protection of them will become much less important compared with protection of them from bad ideas, which would use creativity to destroy all the backups. That is the ultimate extreme of the process, which is already there in the fact that defending against biological weapons involves not biological weapons, but antidotes. That kind of rapid progress is essential to the future of civilization.

1:00:49 **John Horgan** You're just popping open cans of worms all over the place here, David, and we're basically out of time. But I just wanted to make sure that we touched on, at the very end here, your views on our environmental problems, on global warming, the question of sustainability. You're quite critical of the concept of sustainability and also of what you might call, I don't know, environmental alarmism. You recall hearing Paul Ehrlich give one of his gloom and doom speeches decades ago, and you were pretty dismissive because you thought that Ehrlich wasn't anticipating any technological progress that might help us overcome these problems. So just give us a quick picture of your view on our sustainable or not sustainable future.

1:01:56 **David Deutsch** I think it's a great pity that the issue of how to manage the environment has become a political issue, because as a political issue, it has become dogmatic, and the dogmas on all sides are simply false. They contradict the arguments of my book, and what more can I say? On the one side, we have the people who say that the only way of ensuring our survival in the long run is in damping down our impact on the environment. Now, damping down our impact on the environment is itself an impact on the environment. There's no fundamental difference between changes that we cause, changes that it causes, or changes that we cause by trying to undo things that we have done. All those things require knowledge. All of them require technology. All of them are going to give rise to unknown problems in the future. And on the other hand, there are people who try to deny that physics is relevant if it contradicts a political dogma, and that's not true either.

It's rather unfortunate that in the case of global warming, the exact details of how soon this is going to become a major problem depend on supercomputer simulations. That it is going to become a problem eventually doesn't need supercomputer simulations. It's politically important whether the tipping point is likely to come in fifty years or 150 years. The difference between those is enormously important politically, but it's not at all important technologically, it seems to me. In both cases, we need very rapid progress, and we need to assume that the solution is going to come from this rapid scientific and technological progress, and that this won't be the last problem that ever faces us. What strange arrogance it is among the opponents of arrogance in technology to assume that global warming is going to be the last major problem that will ever face our species. That seems to me ridiculous. And the task of technology is not to optimize the entire planet to solve one particular problem that we happen to know about, but to give us the means of, first of all, addressing problems that we do not yet know about, and secondly, the means to recover from disasters that will also inevitably happen when we make mistakes. Both those things, dealing with unforeseen problems and recovering, require knowledge. That's why we need to increase knowledge as fast as we can.

1:05:21 **John Horgan** One final question, and I'm sorry but your answer has to be fairly brief. I just wonder where your optimism comes from. I hope you don't mind my saying, I don't mean this as an insult, but it approaches a kind of faith. And I wonder if that faith has anything to do with a kind of spirituality on your part, a belief in, I don't know, God or something.

1:05:51 **David Deutsch** Well, first of all I deny it. I deny that I have any faith, religious or otherwise, and I deny that I have any spirituality. And I also deny that this optimism is an attribute of me. It's as if you were saying, "What kind of predisposition to multiplicity led you to become a defender of the parallel universes interpretation?" That's not how it happened. The reach comes from the ideas, not from what I want them to say. So I can no more deny the links between the theory of evolution in biology and in, let's say, human ideas, than I can deny that they apply to one particular animal. If the theory of evolution is true, then all animals evolved. And if somebody wants to say all animals except elephants, or, as historically happened, all animals except humans evolved, then that doesn't make sense as an explanation. And what I'm about, what my books are both about, is taking explanations seriously and requiring them to be good explanations, not requiring them to have predetermined implications.

1:07:19 **John Horgan** Well listen, there are a lot of pessimists out there as you know. I'm sure you've gotten some pushback against your vision of the future. So I urge them all to read your book and give your ideas a chance, and it might even make some of those pessimists out there a little less gloomy.

1:07:41 **David Deutsch** Well that's great, and I'm glad that you're not one.

1:07:44 **John Horgan** I'm working on it. Thank you very much David, it was really a pleasure.

1:07:50 **David Deutsch** Okay, nice talking to you.

1:07:52 **John Horgan** All right, same here.

LOGAN CHIPKIN:

CONSTRUCTOR THEORY'S APPLICATION TO INFORMATION THEORY AND BEYOND

About the interviewer: Logan Chipkin is President and Cofounder of Conjecture Institute, a nonprofit dedicated to spreading and applying humanity's deepest ideas.

Series: See the 'Talks' section on the constructor theory homepage: https://www.constructortheory.org/.

About the organization: Conjecture Institute is a registered 501(c)(3) nonprofit organization dedicated to spreading and developing the philosophy of Karl Popper and David Deutsch. We focus on applying critical rationalism to various domains including physics, AI, Taking Children Seriously, aesthetics, and economics. Through research, publications, and educational initiatives, we work to dissolve barriers to progress and foster a deeper understanding of how knowledge grows.

Organization homepage: https://www.conjectureinstitute.com/

EPISODE DETAILS

Date	September 24, 2020
Interviewer	Logan Chipkin
Source	YouTube
Episode	Interview with David Deutsch about Constructor Theory's application to information theory and beyond.
Description	We discuss the constructor theory of information, some of its conjectured principles, the unification of classical and quantum information, and more.
Link	https://www.youtube.com/watch?v=P8bCBypqrEw

Ideas

- Using constructor theory, one can establish the minimum criteria for a physical system to be able to store information, and to be able to process information, which is what makes computation—classical and quantum—possible. One can then use these criteria to distinguish between candidate laws of physics that could be viable (those that allow for information and computation) and those that are not viable (those that do not allow for information and computation).
- The principle of interoperability of information expresses a regularity that all scientists take for granted: that information in one physical medium (say, a book) may be faithfully copied into another medium. All known physical theories have the property that, wherever they instantiate information, this information is interoperable with other information.
- When you have information being exchanged between two well-understood systems via a medium that isn't well understood, you can still make predictions using constructor theory.

Topics Chiara Marletto • constructor theory of information • counterfactual properties of information and computation • dark energy • dark matter • information media • interoperability of information • measurement • physical principles • prevailing conception • quantum information • super information • the interoperability principle • the principle of conservation of energy • the principle of consistency of measurement

Episode title and description reproduced verbatim from the original source.

Transcript

0:14 **Logan Chipkin** All right, I'm here with David Deutsch. David, thanks for joining me.

0:18 **David Deutsch** Thanks for having me.

0:20 **Logan Chipkin** Sure. So we live in a world of classical programmable computers that have been very successful in problem solving all sorts of problems in our civilization, and researchers are also making strides in quantum computing. Given our civilization's success in employing concepts like information and computation, why was a constructor theory of information necessary in the first place?

0:45 **David Deutsch** It arose out of several motivations, initially just from constructor theory itself, because we needed to formulate rigorous concepts of things like measurement and possibility that didn't rely on existing, particular physical theories like quantum theory. So we needed to know what it means in principle for, say, a theory of physics to support information, because if you think of laws of physics, any old logically possible laws of physics, there wouldn't be information in the worlds that they describe. So what does it take to make information? If you were confined to quantum theory, you'd say, "Well, [a system] can store information if it has an observable with at least two distinct states which are measurable." And something like that would do in quantum theory. But then quantum theory [has] a Hermitian operator, and it has orthogonal states, and all those things are rooted in just quantum theory. And we want a thing that will apply to all theories, including ones that haven't been invented yet, all reasonable theories. In fact, this will be usable as a criterion for whether a proposed theory is reasonable.

2:26 **Logan Chipkin** What is it about information that makes the prevailing conception of physics, namely expressing laws of physics in terms of initial conditions and dynamical laws of motion, so difficult to capture, whereas constructor theory, which is all about possible and impossible transformations or tasks, seems much more up to the task, as it were, of explaining and capturing the regularities of information?

2:49 **David Deutsch** Yes, that's exactly why. It's because constructor theory has counterfactual properties at its heart. So possible and impossible are both counterfactual concepts, and information is inherently counterfactual. For example, let's suppose somebody measures a constant of nature, like pi or something. Now, you can't use pi to send or store information, because pi only ever has one value. Whereas whether, for example, a spin of an electron is pointing up or down, [that] can be used to store information. And if you have more than one spin, and they can interact in certain ways, then they can be used not only to store information, but to process information. Using constructor theory, we can establish the minimum criteria for a physical system to be able to store information, and to be able to process information, which is what makes computation—classical and quantum—possible. And then you can say, "Here's our defined class of conceivable laws of physics, which allow information to exist and information processing, which means among other things, it allows measurement and computation, and, therefore, growth of knowledge." And the rest would not allow that. And we won't be looking there for potential laws of physics, because those laws wouldn't allow physics to happen.

5:02 **Logan Chipkin** Following that argument, would you say [that] a similar argument could be made for why it was important to establish concepts like measurement and distinguishing between outcomes of an experiment, which you've touched on, in constructor theoretic terms, because it makes it then easier to look for future object-level laws of physics that must conform to the constructor theoretic definitions of these concepts?

5:28 **David Deutsch** Yes. Of course, it is possible that constructor theory as we conceive of it isn't true. And that somebody invents a theory of physics which violates the principles of constructor theory. So there is no such thing under that theory as, for example, information or measurement or computation, but that nevertheless, the theory is testable in some sense that transcends our present concepts. But even then, that would be useful, because if you had such a theory on the horizon, and it looked as though, for example, it wasn't going to allow measurement, therefore, wasn't going to allow science as we know it to exist, then that would be an early indication, either that the proposed theory isn't true, or if it is true, that it requires a conceptual revolution larger than one would at first thought think, because such a theory might arise by just changing a few innocent-looking parameters in quantum theory. And you might then get a theory that didn't support science. And then you'd have more than just technical difficulties in physics, you'd have profound philosophical difficulties in setting up that theory as well. And constructor theory would have been an early indication of that. But I don't think that's going to happen. I think that future theories are going to conform to constructor theory. But that's just what I think.

7:02 **Logan Chipkin** Time will tell. So let's talk about classical information in particular, which is the world of bits, as most people are familiar with them. What is classical information in constructor theory? And by expressing classical information in constructor theoretic language, what problem or problems have you solved?

7:26 **David Deutsch** The definition of classical information under constructor theory is that a system is an information medium if it can be in at least two different states, [and that] these states can be permuted by possible operations, that is, the operations of arbitrarily permuting those states is physically possible. If that defines an information medium, then you can define information as being present in a system if it is in one of those information states and could have been in the other or in one of the others. So there's your counterfactual definition of information.

Apart from the fact that this works, the only concrete results that we can display at the moment is the fact that this is also enough to define quantum information. Given that framework, you can define quantum information in a very simple way. You can say that an object can hold quantum information if it has some states that can hold information, and if it has another set of states that can also hold information, and if the union of those two sets of states cannot hold information. And when you have a system like that, in constructor theory, we call it superinformation, but the only practical theory at the moment that has superinformation is quantum theory. So our result is that quantum information and all its important properties just follow from this small difference in the way that information appears in the theory. In classical physics, it appears in the way that every classical information medium has only one maximal set of states that can hold information. [We] can have another one, but it won't be compatible. And then quantum theory has two compatible sets of states where the union of the two is not a set of information states. That is, by the way, because they can't be distinguished by any physical process. They can't be reliably distinguished by any physical process. So that's our main result in regard to information to date.

But, as I think Sam told you, what we're hoping for is that this notion of information and therefore of measurement and computation can tell us a measurement theory in new proposed laws of physics, such as qubit field theory, where the existing concepts are not enough. In qubit field theory, we have separated systems where the observables do not commute with each other, and yet causality is maintained. And that's conceptually very difficult. And so you have to build a new theory of measurement. "We're not in Kansas anymore," as we keep saying. And to build a proper theory of measurement in quantum theory took, depending on how you count it, thirty, forty years of head-scratching and thinking that, "This can't be right. This is impossible." We think that constructor theory provides a very powerful tool for setting up a theory of measurement within a newly proposed theory of physics, even if it's very conceptually strange.

12:31 **Logan Chipkin** It reminds me of the fact that you can ignore fundamental principles in a given domain so long as you're far away from the limiting case. By which I mean, it's easy to work in Newtonian physics if you're not approaching the speed of light. And it strikes me that measurement feels similar, where you can rely on your intuitions of how measurement ought to work. But as you said, once you start leaving Kansas, you really have to understand in physical terms what constitutes a measurement. Does that seem reasonable to you?

13:04 **David Deutsch** That's exactly right. And what you just said about measurements applies to information. It's really the same problem. Measurements and information are the same problem. And even in relativity, Einstein had this problem of defining what measurement is essentially. And the concept of measurement had to be different because there's no such thing as simultaneity and so on. That problem was important and he solved it, but it was minor compared with the analogous problem that arose in quantum theory. And I think that future theories are going to challenge our intuition more than previous ones. This has been the experience that new fundamental theories in physics challenge our intuitions in unpredictable ways. And with constructor theory, we hope to have a tool that helps us to formulate our intuitions properly given the new theory, whatever it is.

14:21 **Logan Chipkin** So is it fair to say that with your and Chiara's work on constructor theory of information, you've effectively unified classical and quantum information and it's only that they differ in one property?

14:34 **David Deutsch** Yes. I think that is what this work has done. And I would hold that out, as I just said, as being our one concrete result of the constructor theory of information so far. Maybe Chiara's results in constructor theory of thermodynamics also count, because she uses the constructor theory of information there as well. So maybe her results in thermodynamics count as well. But the real fun is going to begin when we apply constructor theory to physical theories that go beyond quantum theory, that are incompatible with quantum theory. And that's what we mean by not being in Kansas anymore. And I think that is when constructor theory will come into its own. My guess is that it just won't be practically possible to make progress without an overarching framework like constructor theory. Now, as I said, constructor theory could be wrong, but in that case, trying to develop it will show us what is needed instead. I don't think we can make progress without such a framework.

16:10 **Logan Chipkin** Right. And even if constructor theory is wrong, as you say, every error will show us maybe why it's wrong and we can progress in other ways. Either way, it seems like it's a conceptual revolution, as I think you mentioned earlier.

16:24 **David Deutsch** I think so. Yes, I think that's very much the case.

16:28 **Logan Chipkin** So much of constructor theory is about and is expressed in principles, which are laws that constrain other laws. And you've conjectured in your paper with Chiara several principles that explain or capture the regularities of information. So let's just go through a couple examples. First of all, there's the interoperability principle. What regularity of information does that capture?

16:52 **David Deutsch** We take for granted that if we have information in one physical medium, let's say a book, even if it's an ancient book that was produced by a culture that we don't know about, we don't yet know what the symbols mean or what their language was and so on. We take for granted nowadays that the information content of this book, and therefore the knowledge content as well, can be faithfully copied into a different medium. For example, magnetic domains in a microchip or into sound waves when somebody reads it to somebody else, or indeed into our retinas and brains as well and so on. All these are examples of information transcending the substrate in which it's instantiated. So information is this weird hybrid of a thing that information can only exist in physical form, and yet it is independent of the specific physical form in which it is ever instantiated.

What I've just described now, although it's extremely familiar and taken for granted, is not logically necessary at all. It is conceivable that physics could be different and not satisfy this principle, so that you could have some physical objects that have information and have science possible and observables and measurements and even civilization and so on, and that it simply wasn't translatable into another physical system, which could also have science and the whole edifice of knowledge, and yet they weren't intertranslatable. So in theory, aliens from another planet could be based on different physics, not just different physical objects, but different aspects of physical laws, could come and visit us and could be fundamentally unable to communicate with us. This is logically possible, but the principle of constructor theory says that [it] is not so, and all known physical theories have the property that wherever they instantiate information, it is indeed interoperable with other information. So it can be information in electron spins, can be translated into information in microwave cavities, which can be written down and published as ink on paper and the whole works, and all with arbitrarily reliable copying. [This] seems to be a feature of our universe. It is, as far as most people can conceive, as far as I can conceive, a necessary feature for knowledge and for science as we know it to exist at all. And although it is a feature of all known physical theories, there's no known physical law, or there had [not] been before constructor theory, there [was] no known physical law that implies it. It happens to be a feature that laws, as far as we know, obey. But unlike, say, the law of conservation of energy, no one has expressed this regularity as a law until constructor theory came along. And so that's the principle of interoperability of information.

21:19 **Logan Chipkin** And you had mentioned the scientific method in your answer. So I want to talk about one more principle. First of all, it does seem from your work in the constructor theory of information [paper] that you're actually integrating parts of the scientific method itself into fundamental physics. Tell me if I'm wrong, but that's certainly what it seems like.

21:39 **David Deutsch** Yes, I would rather say that we're expressing parts of the scientific methodology in constructor theory. Specific principles like interoperability could be modified without making the rest of constructor theory fall. It's just that it could only be modified in a way that was compatible with the rest of the theory. And if the theory has to be modified too much, then it's a matter of degree whether you call it the same theory. It might be some conceptually different framework from constructor theory. So the interoperability of information is related to the comprehensibility of the universe in this way. If the universe isn't wholly comprehensible, which by the way, philosophically, I think that's ridiculous, but as a physicist, I have to allow it as a possible property that a theory might have. Then in some ways, constructor theory could be altered to reflect that. But, as I said earlier, at least it would give us an early warning that something profound is happening when we postulate, something which has profound implications for physics, for the laws of physics as we know them. The laws of physics [would] have to be formulated not just differently, but with different modes of explanation—not just with different explanations, which have yet to be discovered. Constructor theory would give us an early warning of that. And if these new theories had that property and made sense, then they might be accommodatable in a modified version of constructor theory. But we're being very speculative now. I don't think any of that is going to happen. I think constructor theory will be a reliable guide to, not to what the next theory is, but to what the next theory can't be, or what the next theory can't be without a revolution bigger than it looks at first sight.

24:23 **Logan Chipkin** Right. One of the features of constructor theory is that because it's kind of a theory about theories, it forbids certain kinds of theories from being possible. Whereas in the prevailing conception, the theories forbid just what physical phenomena are possible. So it's a higher level theory in that sense.

24:46 **David Deutsch** Yes. Well, the prevailing conception theories forbid what is possible under that particular theory. So the theories don't speak about each other. The laws of dynamics don't say that the perpetual motion machine is impossible. They say a perpetual motion machine is impossible under Lagrangian dynamics, let's say. But physicists have conjectured for over a hundred years that there is a principle of physics, the principle of conservation of energy, that is a principle about other theories, including ones we don't know yet. And so we use the principle of conservation of energy as a guide to conjecturing new theories. Because we know that if a new theory violates the principle of conservation of energy, then either it's false or we have to reconceptualize the world under that theory. It's not just [a] change of the type of changing the parameter or adding a new term in an equation.

26:07 **Logan Chipkin** Right. And it speaks to the fact that merely falsifying theories is not quite as straightforward as sometimes it's made out to be. There's always background knowledge and so forth.

26:20 **David Deutsch** That's right. And with conservation of energy, this has in fact happened. For a start with neutrinos, that's our favorite example. Neutrinos were discovered because people noticed that energy apparently wasn't being conserved. And therefore some new process must be involved. And this turned out to be weak interactions, weak nuclear interactions. But another example is relativity itself. [The] principle of conservation energy can be expressed in a number of different forms, which in pre-relativistic physics were all equivalent to each other. But it turned out that some of those formulations were incompatible with relativity. Fortunately, there are conceptions of the law of conservation of energy which are compatible with [the] general theory of relativity. Now, it could be that some modification of relativity, like if there was a good theory of dark energy, for example, which violated the principle of conservation of energy, [as] some people have proposed, then the principle of conservation of energy would have been refuted. But this very fact guides our theorizing about dark energy. Because having a theory of dark energy that makes it incompatible with any kind of principle of conservation of energy would be a big revolution, much bigger than it looks.

28:25 **Logan Chipkin** Let's talk about one more principle so that listeners can get a taste of what principles look like in constructor theory, but more specifically in constructor theory of information, the principle of consistency of measurement you expressed in your paper with Chiara. I wonder if you could elaborate on that principle and talk about what regularities you're capturing there that we all sort of take for granted already.

28:52 **David Deutsch** Yes. So this is one of the principles that we use to make sense of information in the context of measurement, which is the main context we originally wanted it in. Now, we assume that this consistency of measurement has to do with information in the sense that when you measure something, say a physical quantity, like the speed of your car, what you're doing is you're causing an information variable to go into a state that represents the physical property of, in this case, the car. So that your speedometer, for example, showing that the needle is at a certain place, that's an information medium in a certain state. And that, according to the theory by which it was designed, represents a physical property of the car. Now, [the] thing is, you can also measure that physical property of the car by a physically very different process. For example, the policeman might be measuring your speed with a radar gun. And the principle of consistency of measurement says that the speed on your speedometer and the speed indicator on the radar gun, if both of those are working properly, will be the same. And if that weren't true, then there wouldn't really be such a thing as measurement in the way we normally conceptualize it.

And again, conversely, if it wasn't true, the situation might be rescuable by a new conception of measurement, but that would be a big thing. It would be a change in our philosophical framework of what science is and so on, that is much bigger than could be expressed by just saying that needles on speedometers and indicators on the display of a radar speed measuring device aren't necessarily equal. That sounds like a very tame thing to say. And it is a tame thing to say in terms of the object-level theories involved. But in terms of the principle involved, it would be a huge change. And again, we think that's not going to happen. We think that the world, in fact, obeys the principle of consistency of measurement.

31:51 **Logan Chipkin** So with these principles that you've conjectured, do you expect constructor theory to solve further problems down the road within information theory itself, whether classical or quantum? And I say that to contrast with what we've been talking about, which is [how] you expect constructor theory of information to help physicists or scientists discover future theories.

32:22 **David Deutsch** Yes. Again, Chiara has been working on this. The constructor theory of information does help with existing theories in situations where, although there is no new physics involved, we don't know what physics is involved. For example, in a situation where systems are interacting, two systems are interacting via another system that is not understood, or which is too complicated to analyze explicitly, then, because of the principles of constructor theory applying, we can say something about that interaction. We can make predictions about that interaction that are independent of the intervening process, provided that the intervening process obeys constructor theory.

You can do the same thing with energy. You can say that regardless of the fuel that the rocket uses, if you have this amount of energy in the fuel, you cannot reach more than a certain height. So you can say that [even] not knowing what form of energy is being used by the rocket.

And similarly, when you have information being exchanged between systems like gravitational field and electromagnetic field, and the states of an electron, and so on under quantum theory, and only parts of the system are understood exactly, and the medium that is transferring the force or whatever it is from one well-understood system to another well-understood system isn't well understood, you can still make predictions using constructor theory. Chiara and Vlatko have several ideas where they have elaborated this into a useful form that may even be usable in real experiments at some point.

34:57 **Logan Chipkin** Since the constructor theory of information principles are expected to be universal, you had mentioned dark matter earlier, or maybe dark energy. Whatever they are, [shouldn't] we expect them to also conform to the constructor theoretic principles of information and the rest of constructor theory's principles?

35:17 **David Deutsch** Yes, in my opinion, it's not on the cards that they won't obey it. Of course, I could be wrong, and as I said, if I am wrong, there would still be useful things to be found from constructor theory. But yes, I think there is no sign in any of the problems arising from either dark matter or dark energy, there is no sign that constructor theoretic principles are being violated.

35:51 **Logan Chipkin** All right, David, well, this has been very interesting and very informative, and I really appreciate your time.

35:57 **David Deutsch** Well, thank you very much. It was enjoyable answering these questions, and I always learn something.

36:06 **Logan Chipkin** Yeah, well, you and me both. Have a good rest of your day.

36:10 **David Deutsch** Same to you.

BRUCE NIELSON:

THEORY OF ANYTHING (WITH SADIA NAEEM)

About the host: Bruce Nielson is interested in intelligence, artificial general intelligence, critical rationalism, the four strands, and epistemology. He delves into all of the above in his Theory of Anything Podcast.

About the organization: The primary purpose of the *Our Karl Popper* blog is to collect stories from from all corners of the world about how the philosopher, Karl Popper, has made a difference in people's lives; i.e. how Popper has acted as a source of insight and inspiration either directly through face-to-face interactions with him or indirectly through his students, research assistants, associates and the large volume of work that he left behind.

Host homepage (Sadia): https://www.sadianaeem.org/
Event: transcontinental Popperian ZOOM meeting Meet 'n Greet
Organization homepage: https://ourkarlpopper.net/about/

EPISODE DETAILS

Date	January 9, 2021
Hosts	Bruce Nielson & Sadia Naeem
Source	YouTube
Show	Theory of Anything
Episode	Theory of Anything Hosts David Deutsch
Description	This was recorded during the 24 hr. transcontinental Popperian ZOOM Meet 'n Greet of January 9th-10th, 2021 organized by OurKarlPopper.net. Bruce was asked to host a session with the subject of David Deutsch and how he brought a whole new generation to Karl Popper's philosophy. But at the last minute, we found out David Deutsch himself was attending. So we redid our plans to allow people to ask him questions.
Link	https://www.youtube.com/watch?v=Q_Cs5iNazB8
Ideas	• Most studies of the foundations of physics falsely conclude that, because of the determinism of the laws of physics and the reversibility of the laws of motion in physics, there is no room for causation. There is, it's just that causation is a high-level concept. • The ought-is distinction merely says that you can't deduce moral knowledge from scientific knowledge, but so what? You can't deduce scientific knowledge from anything. You can't deduce moral knowledge, either. But we're not after deducing knowledge. What we're after is solving problems. And there have to be moral problems as soon as you have a creative entity that is solving problems. Then, moral issues arise because the entity will wonder, "What should I want?" It/he/she has to think about what to want and criticize it and create knowledge about it. • The replication crisis is a small facet of what goes wrong when you apply scientism to psychology and anything that involves human knowledge. If you try to study it as if it were physics, you will be doing scientism, and you will get it wrong.

Topics	animal suffering • artificial general intelligence theories • Bayesianism • beauty • causality • causation • constructor theory • creativity • foundationalism in morality • ignorance • is-ought dichotomy • knowledge-based worldview • no method for problem solving • optimism • Popper • problem situations • prophesying • replication crisis • scientism • static societies • statistics • theory of evolution • twin studies

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **David Deutsch** I've seen people coming on TV and saying how they were inspired by Richard Dawkins and then they say, "Well yes, evolution is the survival of the fittest" and so on. And they just, they haven't got it. And you know, E.O. Wilson hasn't got it, I mean from our point of view. Maybe from his point of view, we haven't got it. From his point of view, Dawkins hasn't got it. So I don't know what the magic thing is that makes progress. If a lot of young people are interested in ideas, then there's going to be progress, even if one doesn't notice it from one's own point of view.

0:49 **Sadia Naeem** And I agree with you, because one of the things I've realized is, it's almost like you have to even go into the psychology of it, too. It isn't just enough for the ideas to be available. If people are not willing, it seems like somehow people are either oblivious or [I don't know] if they're not interested, why they cling to certain things. Sometimes I wonder if they could even just look at themselves, almost turn back on themselves and see why certain thoughts and ideas are coming. I don't know. I really do struggle with that, too. Despite having said that, I think that at least those of us who are willing, who are constantly struggling, it really does help to have those ideas. I mean, we might have gotten there in a while, but most of us, we have limited lifespans, unfortunately, so it helps. Anything helps.

1:45 **Bruce Nielson** I think it takes a while, right? And there's so many ways to phrase things, like even "survival of the fittest." If you think of that as survival of the replicator that replicates the best, you can kind of see how it still fits, right? And so I think that part of it is just [that] it's hard to get away from the memes that exist in a culture. If evolution's about survival of the fittest, you can kind of see how even if you understand Dawkins, that's still true. So you still use that term, even though it's misleading.

2:19 **David Deutsch** Yeah well, Darwin used it. But I don't know, you can't see into people's minds, but I suspect that in many cases, when people say "survival of the fittest," they are imagining animals fighting it out.

2:35 **Bruce Nielson** Yes. I think you're right. I think we have this big mingling in our minds of different ideas, and we don't really differentiate them that well. So I think you're right.

2:47 **David Deutsch** But ideas also have power, and they illuminate people, and there is progress. There really is.

2:54 **Bruce Nielson** Yeah. I agree.

2:57 **Sadia Naeem** It's kind of interesting, too, that when you look into the theory of evolution, of course, they would say that there isn't any directionality in evolution. It's not like things are going towards more complexity. Well, first of all, there isn't a definition of complexity that everybody agrees to. But it's kind of hard to turn away and not recognize that there is something there. We have seen organisms becoming more complex, and it kind of goes hand in hand with the whole thing of recognizing why some people somehow think that there is no progress in ideas.

3:32 **David Deutsch** Yeah. Some people would like to deny it.

3:36 **Sadia Naeem** Sorry. Go ahead.

3:37 **David Deutsch** Sorry. Sorry. Some people would like to deny that there's progress for various reasons: psychological, political, and so on. Once you deny that there's progress, you have a sort of an automatic take on a number of things that you have to be ignorant about if you don't take that view. And so it's kind of comforting. It's kind of pessimism. There's a certain comfort in pessimism.

4:09 **Sadia Naeem** Interestingly, I feel the same thing in evolutionary biology, too. I think sometimes some people have had such a reaction to the whole, because so far many religions have recognized the significance of humans. Like my background, I used to be a Muslim, but we were always told that all the angels bowed down to the human. So God made something and then Satan turned against God…So it seems like a lot of reaction nowadays.

4:51 **David Deutsch** Yeah, like Popper says, "All science begins with mysticism." And I think philosophy began with religion. And "what began with religion" means is that religion was groping towards some truth, attained some truth, some falsehood, and usually tried to suppress criticism.

5:26 **Sadia Naeem** There we go. Sorry about that.

5:27 **David Deutsch** Sorry. Can't hear.

5:33 **Sadia Naeem** Can you hear me, David?

5:34 **David Deutsch** Yeah, I can now.

5:35 **Sadia Naeem** No, I had to mute somebody else that they didn't realize they were unmuted.

5:39 **David Deutsch** I see. Yeah. And I think maybe the atheist movement should give a little ground here and realize that doing better than religion is not synonymous with denying everything that every religion says, because that's like starting from year zero.

6:06 **Sadia Naeem** Yeah, it almost becomes the same sort of thing that you see in different [religions], where people, to give themselves credence, they feel like they have to put somebody else down, because otherwise, how are they going to convince their kids to stick to their religion and not think about something else?

6:23 **David Deutsch** Yes. You're still allowed to deny some aspects or many aspects of the opposing view. But if you try to deny all aspects of the opposing view, you will definitely go wrong.

6:38 **Bruce Nielson** Interesting.

6:40 **Carlos De la Guardia** Reminds me of the Brexit debate. I was rewatching the video with Dominic Cummings, explaining why Leave won the vote. And he said, "Everyone in this room, I guess, predominantly leftist," he was saying, "vastly overvalue the rightness of being on the opposite side of the racists," like Nigel Farage and all these guys. So being on the opposite side of someone who is wrong is not the right way to think about it.

7:10 **David Deutsch** Yeah.

7:16 **Bruce Nielson** All right, just a time check. The official starting of this is in about ten minutes. And we will do a short introduction as soon as it officially starts.

7:29 **Sadia Naeem** Oh and Bruce, I want to let you know, I asked Margaret. We can all record and I actually kind of started recording. Actually, I'm not sure if you can record.

7:42 **Bruce Nielson** I just started recording, too.

7:44 **Sadia Naeem** Yeah, so sounded like she didn't think that it was going to make that much of a difference. But it's up to you, you can go ahead.

7:50 **Bruce Nielson** No, that's fine. We're good. We're good.

7:52 **Sadia Naeem** I'll keep an eye if somebody entered unmuted, then I can kind of maneuver around. So yeah, because we have no idea how many people are going to turn up.

8:05 **Bruce Nielson** Yeah, I have noticed that people are coming into the room unmuted, which is unfortunate. We may have to mute people as they come in.

8:13 **Cameo Duran** Yes, so much for my theory, right?

8:15 **Bruce Nielson** Yeah. And Sadia, you are the only one who can do it. So you're going to have to probably mute people as they come in.

8:23 **Sadia Naeem** All right.

8:28 **Carlos de la Guardia** By the way, David, I have a somewhat random question for you as long as you're here. Did you have any expectations about what would happen when you first published *The Beginning of Infinity*?

8:41 **David Deutsch** Well, I was hoping that people would buy it.

9:01 **Sadia Naeem** Sorry, give me one second. I'm trying to identify. All right, there we go.

9:05 **Bruce Nielson** Sadia, to be able to control that, open up the participants and stretch it out so you can kind of see everybody at once and they'll be near the top talking and you'll be able to mute them quicker as they come in. Sorry. Go ahead.

9:23 **David Deutsch** Yeah well, one thing I thought at the time with [*The*] *Beginning of Infinity*, I ended up finishing it under a deadline, and it wasn't as polished as I was hoping it would be. And I had to leave out an entire chapter that I had planned. It took almost ten years to write, as did [*The*] *Fabric of Reality*. But with *The Fabric of Reality*, I finished it in my own time, and [*The*] *Beginning of Infinity* was a bit rushed. And so I was thinking that it wasn't as good. And although many people criticize it in many ways, few people said it wasn't as good. So, you know, go figure.

10:19 **Bruce Nielson** [*The*] *Beginning of Infinity* actually seems to be the more popular of the two books from what I've seen. Personally, I'm a *Fabric of Reality* fan. I read [*The*] *Fabric of Reality* two years before [*The*] *Beginning of Infinity* came out. So I was anxious when it came out. I'm curious, what was the chapter that you didn't get to do?

10:40 **David Deutsch** I don't know what it would have been called, but it was about scientism and related issues. A few paragraphs of that chapter got into the chapter on choices. You know, the working out [of] how many people go into the museum and come out, and then you form the theory that people are being spontaneously created and destroyed and that idea. That was from the other chapter, but I had been planning a long chapter on scientism. I now think that scientism deserves a whole book, and I am not the person to write it. So maybe that never would have been written.

11:22 **Sadia Naeem** It's interesting you say that because my first experience when I broke away, I don't want to say broke away from religion. For me, it was a very natural progression when I recognized one day that I was an atheist. But I felt a little bit of an isolation in my own community because I was just so weird in that way. But I started looking for other places and there were a lot of atheist groups and free thinkers. And when I joined them, I almost felt like I was going to some sort of a religious place. I really wanted to be with people where I could just literally talk without saying, "Oh, you're not allowed to ask this question." But I didn't find that. That made me realize, when I heard you talk about scientism [or] I read, that clicked right away that, unfortunately, either you have that or the other end where you're just not allowed to ask certain questions.

12:30 **Bruce Nielson** The shutting down of criticism.

12:33 **David Deutsch** Yes.

12:37 **Bruce Nielson** All right. Time check. We've got six minutes before the official start. We've got quite the crowd. I think this might be one of the larger crowds.

12:44 **Carlos De la Guardia** I'm going to say, David, it's been fun meeting some new people. I'm currently visiting Austin, Texas right now. So I know you've got a little bit of history there. And it's funny to see how, or maybe funny is the wrong word, but it's very interesting to note how the knowledge-based view of the world changes the discussion of the whole shape of certain kind of discussions that otherwise would be maybe people-focused, like classes of people and scientists up here and all these sorts of things or just asking questions about where knowledge is created, where conflicts are happening, where disagreements are happening, simplifies so many things to the point where people will ask…my favorite recent thing is that somebody will ask me for relationship advice or something. And I'll give them the same caveat that you always do: "I don't know that much about relationships, but what's the problem?" And then you can kind of just ask a few questions and see, "Okay, well, I can think a little bit about disagreements," and I'm constantly surprised that there's always something to be said. It may not be incredibly relevant, but what my friend told me, and I didn't really expect this would happen, is he said, "Whenever I [talk] to Carlos, and I always tell them, you're effectively talking to David indirectly."

14:13 **Bruce Nielson** You run your David module.

14:18 **Carlos de la Guardia** Yeah. He says, "The problem is unchanged. And yet I feel so much better." And the analogy that I gave him was that he was like someone who had to build a spaceship. And he was currently in the desert. And he had just been transported to a beautiful high-tech facility with all sorts of tools around. He hasn't built the spaceship yet, but suddenly the situation surrounding the problem is now totally different. Whereas, it might have been: "This person doesn't like me." It becomes just about: What knowledge is lacking? What discussion do I need to have? How can I take this person who I thought might disagree with me and that could be a problem and who I might try to lie to or otherwise try to get something and say, "Well, how can I just make the problem an objective thing we can both try to solve and double our efforts and the creative possibilities here?" And so he just seems to have that view that things become so much easier once you have this view of knowledge, even if you haven't directly solved the problem.

15:29 **David Deutsch** Maybe you're describing the transition to optimism. If you think about what's going wrong in terms of a lack of knowledge, although you still don't know what that knowledge is, in a certain sense you know that what's standing between you and the good outcome is a lack of knowledge, and you need to create knowledge. And that already puts an optimistic spin on things even before you solve anything. Whereas if you think of things in terms of people, then everything becomes, "Who? Whom?" the famous thing [that] Lenin is supposed to have said. Just a very accurate description of a whole class of worldviews, "Who? Whom?" And you've got to get rid of "Who? Whom?" If you get rid of it in politics, that's like getting rid of, "Who should rule?" and so on. And presumably, from what you've just said in relationship things, you get away from "Who? Whom?" and you turn towards, "What actually is the problem?"

16:46 **Bruce Nielson** All right. It is time. So, Sadia, why don't we go ahead and start the official meeting, if you could do the introduction and then we will go on from there.

16:57 **Sadia Naeem** All right. First of all, I want to welcome everybody. It's kind of nice to see, I think this has been our largest session. Wanted to start by introducing, I think most of us probably know David, but I wanted to just give a little bit of an intro to David by pointing out that, first of all, the Age of Enlightenment saw the rise of what some of us recognize as Popperian tradition. While Popper's work may not be that well-known, the tradition he presented or talked about has been with us for a few centuries now. I think that David Deutsch has done a wonderful job at bringing Popper's work to people through his books. And he's also encouraged a culture of sharing and criticism of ideas. And most of all, by making himself accessible to social media. I think he's pretty accessible, as some of us know, to Twitter, which is not such a common thing among people who are specialists in certain fields. His work, optimism, and interactions have inspired a culture whereby people have come together to take Popper's ideas to a new level. This has led to groups such as the Four Strands, which Bruce and others contribute [to]. This is a thing where I have to be the moderator. I mean, I have to unmute everybody too.

But this has led to groups such as Four Strands run by Bruce Nielson and others. Individuals have been inspired to create podcasts, YouTube videos, and websites to encourage an open-ended growth of knowledge. David's book opened me up to the ideas that impacted my life in more ways than I can mention in a few words. My primary interest is in foundations of physics, and I'm also an educator. I find myself starting my physics class every year now for the last couple of years with a discussion [of] David's TED Talk on good explanations. And it's interesting how that leads to all sorts of interesting discussions throughout the school year as we look into physics and just overall the connection of physics and reality. We talk about reality. We even talk about multiverse sometimes, and it really gets kids into it. So I really want to thank David for making himself available. And thanks a lot, David, for coming today. And I'm going to hand it over to Bruce now and let him help introduce.

19:32 **Bruce Nielson** All right. Thank you, Sadia. I'm Bruce Nielson, and I think I was asked to help host this session because my experience is fairly typical of probably a lot of yours. So back in 2009, I was a religious blogger, and I had fellow religious bloggers suggest to me to read David Deutsch's book, [*The*] *Fabric of Reality*. And I was enthralled with it. So I started actually blogging about it and things like that. And I spent years actually trying to refute what was in his book and ended up reading a whole lot of different books that were related subjects and eventually became very convinced of all four of the strands he mentions in [*The*] *Fabric of Reality* because of my inability to refute them, inability to find good criticism of them that he hadn't already responded to. And so eventually this even led to me going back to school. I wanted to study this more deeply. I wanted to go back to school and get a master's degree in computer science, study computational theory and related subjects. And so this is something that really has ended up impacting my life quite a bit, just in a lot of ways, starting off as a hobby and then eventually now maybe even turning into a career change. I've started a podcast, [The] Theory of Anything podcast, which is loosely based on David Deutsch's four strands. And Sadia mentioned the Four Strands blog, fourstrands.org. I am the one behind that, that runs that and hosts that. And my co-host, Cameo, is also here. Cameo, do you want to do a quick introduction?

21:17 **Cameo Duran** Yes, I do. Hi, I'm Cameo Duran and I'm Bruce's co-host on The Theory of Anything podcast. And everything I know about David Deutsch came from my involvement with Bruce. I think one of the first times Bruce and I had a conversation, it really quickly veered into the Popperland and his passion around the four strands and was the primary reason we started the podcast together, just because we really enjoyed discussing knowledge and that's why I'm here. Hi.

21:54 **Bruce Nielson** Thank you. So what we're gonna do for this session, I do have some questions if people run out of questions, but I wanna give people a chance to actually just talk to David and to ask him questions and to pick his brain and things like that. So the way to do that so that we don't get things out of control, maybe put a question in the chat and Sadia will unmute you and we will just let people have a chance to kind of talk with David and have fireside chat here.

22:27 **Sadia Naeem** You could try that option of the raise hand. If I see a [raised] hand, I'm gonna try my best to spot you and then this way nobody goes over each other.

22:36 **Bruce Nielson** Yeah, raise your hand. That's a good idea.

22:38 **Sadia Naeem** Yeah, so it seems like we have one there. Please go ahead. Ernst, is it, please? I'm sorry if I mispronounced.

22:44 **Bruce Nielson** And I should tell everybody this is being recorded, and I know for sure that we're gonna be releasing this on The Theory of Anything podcast. It probably will get released other places, too. So just be aware of that as you speak up here.

23:00 **Ernst** Thank you. Thank you for hosting this wonderful thing. I was here earlier also. It's been great. So I was thinking a little bit about the transition that you write about, your explanation of why humankind was stuck in static societies has to do with irrational memes. But if that is because you need that, the conception of anti-rational memes, because that's the explanation for why this exponential growth didn't happen because you don't need to assume so much for it to happen. You just make people make small changes into their ideas, and then that will lead to exponential growth. Why wasn't the static society, why didn't it get stuck completely? If this is the question, before the Enlightenment, that the argument is something like: the Enlightenment could have happened earlier, but the Enlightenment happened in a particular culture, and wasn't that culture different than the static culture that preceded it? Something like this.

24:29 **David Deutsch** Yeah, there is a thing which maybe isn't clear in my presentations. When we think about the Enlightenment as distinct from what happened before, there's a selection effect that we tend to think that what happened before was like the Enlightenment except with static societies. But the thing is, long-lived static societies are rare—not as rare as the Enlightenment society, but still quite rare. Most societies, most cultures that have ever existed have not survived very long at all. Back in prehistory, it may have happened more often than a static culture evolving was simply a culture evolving, which did change and then destroyed itself because its changes were not in the direction that would stabilize it. For example, they wouldn't have had traditions of criticism. So maybe they were changing and, as a result, they made many mistakes and, as a result, they were unable to correct them. And so they were killed by the neighboring tribe or they ran out of food and didn't know what to do or whatever. So the sort of natural state of nature, if you can use that concept with humans, you can't really, but the state that humans or prehumans were in when creativity first evolved was maybe better described as just continual chaos and failure rather than staticity. And then staticity sort of emerged out of that sometimes. But because staticity made the cultures last longer and grow more, those are the ones that we kind of see when we look back. We see the ancient Egyptian empire and that kind of thing. And we don't see the many failures that must have outnumbered that culture. I don't know if that answers your question. Maybe I'm missing something.

27:10 **Ernst** Yeah, that was not what I got from the book so far. So it was a little bit different, yeah.

27:20 **Sadia Naeem** First of all, before I tell the person who's next in line, are there any comments to this or any further things somebody wants to ask or add to this particular question? Okay, so I'm gonna go to the next person and I apologize ahead of time. Your name is Pavan, I'm sorry, it starts with a P. Could you please tell us what your name is and please go ahead.

27:47 **Pavan** Hi, David and everyone. My name is Pavan and thank you so much for coming. I think the reason I was being recommended to your book is that I was asking a person on how to do research, and then he recommended this book to me. From his understanding, the most important take from this book is, maybe just for me, but it's about self error correction. So the first question is: Do you have any advice to how this kind of self error correction can take place? Is there a set of questions that a person can question himself in his everyday life, for example, or in his own research? So this is my first question. My second question is, I think in the end, we as [humans], in all sciences, what we're trying to really understand is about causality. But the problem to me is that I'm not a theoretical physicist and I get a sense that my understanding of causality can be very naive and far from what actually is being considered as causality in physics, like the spacetime causality, for example, in physics. But I think at times I can see why things are not causal and are merely correlations, but I find it really hard to give a definition of what is causality. In this scenario, how can I, as a researcher, [try] to probe into these causal relations? This is my second question. My third question, and this is the last question, so right now I am a grad student working in statistics, and I think fundamentally it's a problem of induction that we're trying to combat in our everyday life. So my question to you is: What do you think is the most important thing to do for statisticians or a statistics researcher to help in the process of scientific discovery, what do you think will be the most important thing to [do] for a statistician in the next fifty years, or thirty to fifty?

31:16 **David Deutsch** Well, one can't prophesy, of course. To answer the last question first, because I think that's the easiest. Statistics is an interesting and useful branch of mathematics, and the way it enters into science is that it enters only in what I call in my book the perspiration phase. That is, it is the last step in discovery. It is the part that is not creative, but mechanical. So if we have a mass of statistics and use statistical theory to get an answer out of that, the answer was really created before the data were even collected. And that part doesn't involve statistics. So doing statistics, one has to understand that this is a branch of mathematics, and that it has nothing to do with creativity. Some people think that creativity is just extracting knowledge out of data, but that is the opposite of what the truth is, as Popper has taught us.

32:38 **Pavan** Right, so I can see from your book that it can be kind of used as a tool to reject [hypotheses]. Do you think it's likely that it can be used to discover [hypotheses] as well, or maybe not so much?

32:59 **David Deutsch** I think that's fundamentally impossible for the same reason that any piece of mathematics can't lead to discovery. The piece of mathematics isn't about the world unless you first have a theory that connects it to the world. Now, as I say, speaking of Popper, that leads me to your first question. There is a very nice transcript on the internet somewhere of the lectures, or some of the lectures, that Popper gave to his scientific method class in the LSE when he first joined the LSE. And the first lecture, I think it's the first lecture, begins with him saying, "I think I'm the only professor of scientific method in the British Empire. And the first thing I want to say about this is that there is no such thing as [a] scientific method." And I think this applies equally well to other aspects of Popper's philosophy. There is no such thing as [a] philosophical method or a self-improvement method or a psychological method. It's all opportunistic. It's opportunistic problem solving. So when you said maybe the theme of my book is all problem solving and maybe the theme of all Popper's books is also problem solving—the thing is, there is no method for that. There are various methods for avoiding doing that. And it's a good thing to try to escape from those methods if they are in one's culture or in one's psychology. But that by itself doesn't do anything positive. It merely frees one from the sabotage of those methods.

35:31 **Pavan** It's all about creativity.

35:34 **David Deutsch** Yes. And now what was your second question? I remember the third and first.

35:40 **Pavan** The second question is about causality.

35:43 **David Deutsch** Oh, yes. Well, in what you would find if you looked at modern physics and modern philosophy, what they say about causality is that they basically deny that there is such a thing. Most studies of the foundations of physics conclude—falsely, I think—that because of the determinism of the laws of physics and the block universe and the block multiverse and whatever, and because of the reversibility of the laws of motion in physics, that they equally well predict the past from the future as well as the future from the past or almost anything from almost anything else, that there is no room for causation in that picture. And I think that there is. It's just that causation is a high-level concept. There's no mention of difference between liquids and solids or backwards and forwards in time, either, in fundamental physics. And yet there are well-developed physical theories of both of those things. And causality hasn't really been important in physics for maybe the last couple of hundred years, but that's not really very important. In constructor theory, if I can plug that for a moment, it's much easier to frame a theory of, or frame explanations, in terms of causes than it is in the prevailing mode of explanation. And in other fields than physics, causation is important. Attempts to eliminate causation and try to pretend that one can explain things like human behavior in a deterministic way are dead ends or worse. So that's my answer to the second question.

38:08 **Pavan** Thanks. David, do you mind if I ask you just an additional question?

38:15 **Sadia Naeem** Actually, could I pause? I'm sorry. Could I interject for a second? Because I'm actually seeing a few hand raises. So how [about] we do this? Because I want everybody to have an opportunity to ask. So if you have more than one question, then how about we ask it again and then I will put you in.

38:33 **Bruce Nielson** Raise your hand again if you've got a second question. We'll keep coming back to you as long as there's time.

38:38 **Sadia Naeem** It seems like now we have actually a list of people. So, Ella, could I please let me go over to you next?

38:46 **Ella** Yeah, sure. Can you guys hear me okay?

38:49 **Sadia Naeem** Yeah.

38:50 **Bruce Nielson** A little bit soft, but I can hear you, Ella. Ella, you're muted now.

39:00 **Sadia Naeem** For some reason we can't hear you, even though you're not muted.

39:05 **Ella** Next person and get me afterwards.

39:07 **Sadia Naeem** I can hear you now.

39:08 **Bruce Nielson** You're back now.

39:10 **Ella** Okay, hopefully this will work. So, David, I'm very interested in artificial general intelligence, which is to say I'm interested in trying to understand the mind and the way that the mind creates knowledge at a level of detail that is sufficient that we can implement it on a computer. And so my question is about the logic of how minds manage to create knowledge and the extent to which it's similar or different to biological evolution and the way that knowledge is created there. So my question is: Do you think that replicators are involved in the way that minds manage to create knowledge? I think in biological evolution, we know from Darwin and Dawkins' theory that the replicators are sort of the key explanation for why biological evolution manages to create knowledge. And so I'm interested in whether you think that there's something similar going on in the human mind, some sort of pool of competing replicators, or do you think that there's some other process that's responsible for creating knowledge in human minds?

40:09 **David Deutsch** To some extent, that's a question of implementation. I don't know how creativity works in the human mind. If I knew, I'd be really working hard on that now, if I had any kind of idea that I thought was halfway viable. In regard to replicators, my guess is that that's not how the implementation works in the mind. There could be a logically equivalent implementation in terms of replicators. But the thing is, in the mind or in a computer, you could save memory space just by, rather than by having multiple copies, you just have one copy with a number. There are 10,000 of these, which is a bit like saying this thing is worth 10,000 of this other thing which hasn't done the equivalent of replication. I should say, as I say in the book as well, that I don't think we understand biological evolution well enough, either. Maybe one route towards AGI would be to do the equivalent of artificial biological evolution first. It may or may not be a good route. Replicating a bird's wing was not the best route to artificial flight. Although the underlying theory is the route towards it, the underlying theory of how a bird's wing works is the way to make an airplane. So I doubt that there are replicators in the brain.

42:18 **Ella** Great. Thank you so much.

42:21 **Sadia Naeem** All right. Any comments on that? All right. Okay. Vaden, you can please go ahead.

42:33 **Vaden** Hi, I'm a PhD student at UBC in machine learning. I keep trying to get people to think about knowledge in my community, and they keep confusing it with information. I have a very difficult time explaining to them that I'm trying to refer to something else without just pointing them to your books and Popper's books. I guess I'm curious to know how you think about the difference between knowledge and information, and then also if you have any communication strategies that you could offer in terms of how to get people to realize that I'm trying to talk about something that's not information when I say the word 'knowledge.' Thanks.

43:12 **David Deutsch** Yeah, the only communication strategy that works apart from spending many years writing a book is conversation. And you just get together with someone and try to overlap your problem situation and then something happens. I think of knowledge as a species of information. And I've at various times used several different characterizations of what makes it different from other information. And my most recent choice is to say that knowledge is information with causal properties. There's causation arising again. So knowledge is that property of a computer program that makes it do something useful. For example, you have a word processor and the word processor is useful because it knows, the programmer, of course, is who generated the knowledge, but the programmer has put into the program knowledge of things like: there are such things as words, there are such things as letters and sentences, there is such a thing as correct spelling and incorrect spelling, and so on. And there are different aspects of the context which have to be taken into account and so on. So knowledge is information with causal power.

Also an interesting thing about it, both knowledge and information are very unusual, they're abstractions, and many people don't like to believe that abstractions even exist. So that's something you have to persuade them [of]. But then, further, information and knowledge are extremely unusual abstractions because they only exist when they're physically instantiated, and that's another confusing concept that I sometimes have to work hard to persuade people, or rather to get people to see what I'm even talking about, whether they agree or not, what I'm even saying. So I don't know that I have anything better to say about how to persuade people of things. I don't know that it's even a good idea to try to persuade people of things. What's more important is to have an interesting discussion.

46:23 **Vaden** [I'm glad to] hear that you struggle with it, too.

46:27 **Sadia Naeem** All right, Clovis, would you like to go ahead please?

46:33 **Clovis** Yes, thank you, David for doing this. My question is about moral philosophy and moral truth. This is a topic you've touched before. I'm concerned about how the is-ought dichotomy is interpreted as often hopelessly nihilistic or that it condemns us to relativism and the idea that if moral values can't be derived from facts, they can't be true because they don't refer to objective entities. So for many people who believe in moral truth, the dichotomy is often perceived as a deep problem and a deep mystery. And to me, that seems to be an error because the impossibility of deducing values from facts does not amount to a demonstration that they're false. It's not a refutation. And in a sense, moral ideas [can] be refuted by mere facts any more than they can derive from them. And I find myself in the minority of people who believe the is-ought dichotomy is true but who also believe that it doesn't keep us from creating moral knowledge. And Popper described the position that he called critical dualism that I interpret in this way. And so my question is this. I know that you've talked about the fact that morality is a form of knowledge. And I wanted to ask you: How do you understand the is-ought dichotomy? How does it bear on your concept of moral truth? And does the concept of truth apply to moral propositions? Thank you.

48:02 **David Deutsch** Well, my opinion is it definitely does. And I agree with everything you said there. So that's basically my position as well. Popper, it's a bit hard to interpret on issues of objective morality because he doesn't really discuss that point. You can only infer Popper's position, as far as I know, anyway. I haven't read everything he wrote. You can infer when he says, for example, that we can make moral progress, and also that there is such a thing as making progress in philosophy generally, that he certainly rejects the position that science is the only thing one can make progress in. I like to use the argument that, when people say that there's a difference between the possibility of progress in morality and in science, in that in science, we have this method of experiment that can take us forward and in philosophy, we don't. Well, I think that's an un-Popperian point of view because that's more like the Duhem-Quine view. It's a bit arbitrary to say that scientific knowledge is possible, if at the same time you're going to take that critique of moral knowledge seriously because the same critique that the deniers of moral knowledge take seriously has been used by many people to deny that scientific knowledge is possible and all knowledge is conjectural. The fact that you can't deduce it from anything is irrelevant in all fields. Knowledge can never be deduced.

So the ought-is distinction merely says that you can't deduce moral knowledge from scientific knowledge, but so what? You can't deduce scientific knowledge from anything. So you can't deduce moral knowledge, but we're not after deducing knowledge. What we're after is solving problems. And there have to be moral problems as soon as you have a creative entity that is solving problems, then the moral issues arise because you've got to wonder, "What should I want?" When you're wondering, "What should I do next?" you can't gaze into your navel and find what you want about everything. You've got to think about what you want and criticize it and create knowledge about it. So I think one can take a completely uniform view of all those fields, and therefore the ought-is distinction is not epistemologically relevant. It's not relevant to what kind of knowledge we can create.

51:20 **Sadia Naeem** Actually, I had a question which was similar to that. If you guys don't mind me interjecting in there. My question was about, usually when I talk to people about that, one of the [questions] that's raised is that, when it comes to science, the laws of nature constrain everything. Like we don't have a choice in that, but the moral seems to be different. I guess one of the [differences] between morality is that even if we claim that we discover moral principles, then we still have a choice. We are not bound. It's as if they feel like there is something more concrete in science. Would you like to say something about perhaps… you have any ideas about roots of morality in the sense of, do you tie it to? I've listened to your discussion with Sam Harris. It doesn't seem like you tie it to anything to do with neuroscience, but do you think about it? Is there something at the back of your mind as to what are the roots of morality?

52:24 **David Deutsch** I think in general, it's not very helpful to think about what the roots of something are, because when you find some roots, there's always going to be roots beneath that, and you'll never get to the bottom of it. So foundations are sometimes useful, but not because they're underlying everything, but because they reveal something of the structure of things. I'm a theoretical physicist, I work on the foundations of physics. When you make a terrible mistake at the foundations of physics, you may get ridiculed and you may lose your income and so on. But when you make a mistake at the foundations of morality, the physical world will come for you much worse. I'm not only talking about other people coming for you. Even if you were a person on a desert island who made moral mistakes, it would cause physical trouble for you. You would make mistakes in your life, which might shorten it just from making a mistake in morality.

So I don't think this distinction that morality is a matter of choice is true, or at least it's no more a matter of choice than any other ideas are a matter of choice. We choose and create our own ideas according to our values about what's true. But our values about what's true, even though they are completely changeable, are not at all arbitrary. Maybe the best example of this is pure mathematics. Some people are reduced to claiming that mathematics is arbitrary. Really, mathematics is just the study of what mathematicians think it is clever or glorious or whatever to think about, which reduces mathematics to basically a study of human brains, mathematicians' brains, or the brains of a community of mathematicians. But it's simply not true. Mathematics is the study of abstractions that actually exist and properties of them that exist and are independent of us. We can choose which mathematical objects we think are interesting and worth trying to understand, but we can be mistaken and we can follow dead ends. I think in mathematics it's also unusual to run into a brick wall like that.

By the way, I think that running into a dead end and making large mistakes, unless they kill you, it's not all bad. In fact, it can be just as good as successfully discovering things, which the latter can leave you feeling empty. Whereas, as Popper says, if you're engaged with problems, even if you never solve them, then you're still having fun.

56:04 **Bruce Nielson** I have one quick question. I enjoyed reading your constructor theory paper. You made a very big deal in that paper, though, about it underpinning the rest of physics. I kept wondering why that was, because it seemed like it would be a valid theory about constructors in the same way that information theory is a valid theory about information, or computational theory is a valid theory about computation, without the claim that it underpins all of physics. What was the motivation there to say that? Is that an absolutely necessary motivation, or would it still be a good theory without that?

56:41 **David Deutsch** Well, I guess that no particular motivation is ever essential. I think constructor theory could stand on by itself, but, rather like philosophy, if there were no applications to anything else, then it would be useless. It would just be a piece of mathematics. The reason I think it's important that it underlies many areas of physics is just that I think it does underlie them. I think that there are several areas of physics where progress has been stalled because of the assumption that the prevailing mode of explanation, namely initial conditions and laws of motion, is the only legitimate form. Without ever being stated explicitly, it's taken for granted that a valid explanation in physics has to be of that form. And yet, already in existing physics, there are explanations which are of the constructor theoretic form instead, and cannot be expressed in terms of initial conditions [and] the laws of motion. That is kind of shrugged off because people think it's not legitimate. In thermodynamics, there are explanations that seem to directly conflict with explanations in terms of initial conditions and laws of motion.

The conventional response to that is basically to say, "Well, thermodynamics isn't really true. It's just an approximation scheme, and at root, these quantities like work and heat and the laws of thermodynamics are not actually true." But that's just a prejudice. My feeling is that in that area and in many other areas, such as theory of computation, and in areas of physics where initial conditions and laws of motion approach has been successful, I think in all those areas, there is scope for making progress via constructor theory, if constructor theory is true. And probably not if it isn't. And we'll find out if it's true only by trying to make such progress using it.

59:25 **Bruce Nielson** I wanted to give Dwarkesh a chance to ask a question. He wasn't able to, through his interface, raise his hand, and he did it about this point. So are you still there? And can you ask your question?

59:35 **Dwarkesh Patel** Yeah, I'm here. Thanks. Hey, David, I'm a big fan. I just wanted to ask you, this [isn't] my view, but I just want to play devil's advocate here, because I don't have a good rebuttal to this argument, which is: There's a Bayesian critique of Popper, which is that verification and disconfirmation both reveal information about a theory, and that while Popper can deal with disconfirmation, there's no way to integrate evidence that verifies a theory. Bayes is backwards-compatible with Popper, in that it can integrate verifying and disconfirming evidence. It just weighs disconfirming evidence higher and updates heavier based on that. So how would you deal with that criticism?

1:00:15 **David Deutsch** I think the context in which that criticism arises contains mistakes. First of all, the context is that there is some data or information, which we receive, and then we have to make sense of it, either by refuting a theory or by confirming a theory or whatever, but we start off with data. That just isn't true, as we have learned from Popper. So in that respect, the whole picture of science, and of thinking generally, that underlies that critique is just wrong. So that's, like, where science is coming from.

Then there's where science is going to. So this critique suggests that what we're trying to do, where science is going to, is getting justified beliefs. [That] what we really want to do is to make the probability that we assign, or the credence that we have, for true theories should go up. We need some method that will make the credence of true theories go up. And then they say, "Well first of all, Popper seems to only have a method that makes credences go down. So, how can that possibly be a picture of science?" Well, the answer is that science, from beginning to end, doesn't resemble that picture. So, science is problem-based, and the way it proceeds is by conjecture. And after it has problems and conjectures, it has criticisms. And none of that appears in the Bayesian picture. So of course they're going to think that [the] Popperian view of science doesn't adequately represent science.

But what has really happened is that their picture of science, which is basically empiricism, inductivism, some kind of that, is just wrong root and branch, false root and branch.

1:02:49 **Sadia Naeem** All right. Thank you. Mike.

1:02:54 **Mike** Yes. Hi, everyone. Hi, David. So I was wanting to ask you about modes of explanation and knowing how important they are to kind of structuring some of your work. And Bruce just brought up constructor theory, which I think you might describe as its own mode of explanation. I was trying to particularly link it to computation. So I have your shorthand, "If you can't program it, you haven't understood it." I was wondering if you follow that, is inventing a new mode of explanation, is that synonymous with inventing a new type of algorithm? Is the link to computation and explanation, can it be forged in that way? But not yet. You don't have to speak specifically just to that.

1:03:44 **Mike** I'm sorry, David, I think you're muted.

1:03:45 **Sadia Naeem** David, I think you're muted.

1:03:50 **David Deutsch** Sometimes the space bar works and sometimes it doesn't work. Okay, I'm pressing it down quite hard. Can you hear me?

1:03:58 **Bruce Nielson** We can hear you now.

1:04:05 **David Deutsch** Yeah, yeah. I'm reluctant to reduce things to algorithms. I think that usually sucks the creativity out of the picture and makes it wrong. I'm trying to think whether this maxim, "if you can't program it, you haven't understood it," which is really a bit of a paraphrase of Feynman, whether this applies to everything or just theories about how information works in the world, and in particular, AGI and so on. So if you can't program an algorithm, you haven't understood it. If you can't program any kind of information process, then you haven't understood it. If you can't…Say you have a process of how stars work, a theory about how stars work…I'm thinking out loud here. Then it's also true that if you can't program that, you haven't understood it. But that doesn't mean programming the motion of every molecule in the star. It means programming the things that the features of the theory, of your explanatory theory, that your theory says explain the star. So it's those that you have to be able to program, but finding out what those are is not a matter of programming anything. It's a matter of creativity and problem solving. So my tentative answer is: That maxim doesn't apply to everything. It doesn't apply to creating the knowledge to do that.

1:06:16 **Sadia Naeem** And Mark, would you like to go next?

1:06:22 **Mark** Yes, you can hear me? So thanks for doing this so much. It's really an honor to talk to you. But I find that all the things that we can assign objectivity to in life, I feel like the hardest one for me personally is aesthetics. So for instance, I find the cave paintings of Altamira and Lascaux to be beautiful, but the reason I do is because of how old they are. It's humankind speaking to us from 30,000 years ago trying to survive the harsh ice ages. And I feel like if someone painted the State Rotunda the same way with the bison and everything said it was a masterpiece, I'd probably want to slap them in the face and say, "I don't find that very beautiful." So I don't know if [it's just] me ascribing aesthetic value to the cave paintings of Lascaux because of the romantic notion of humankind painting them so many years ago, and then what are they trying to say, if they're trying to say anything else at all. Is it fair to describe the aesthetic value to that, for reasons like that, or should we just judge it for just how it looks and it shouldn't be the environment who did it and what they're trying to say? If that makes sense.

1:07:39 **David Deutsch** Yes, I think to some extent this is just a matter of the fact that language and terminology aren't, we don't have an absolutely exact language to describe everything we want to talk about. So often we use metaphors and often we use terminology that slides over from one area to an adjacent area and so on. So a mathematician can describe an equation as beautiful. A person can describe someone's mind as beautiful and they mean something by that. They mean something objective by that, but it is not the same thing as what we mean when we describe, say, a piece of music as beautiful or a sculpture as beautiful, and even with those things we may describe a painting as beautiful because it is very apt in a certain situation. Like, I don't know how you judge Goya's painting of some partisans getting shot. How do you separate the beauty of the fact that he's captured by the way a very ugly situation? How do you separate that from beauty in the sense that if the same skill and insight had been used to describe an orange harvesting festival? It could also describe that as beautiful, but there'd be a different kind of beauty being described there.

I think there is such a thing as artistic beauty, which is often mixed with other values that we want to put into an object, and maybe we shouldn't get hung up on whether that is really beauty. [That's] kind of essentialism to ask that. The thing is that there are many features of an object that are desirable, and the cave paintings are desirable in one sense and are clearly rubbish in another sense, and there's nothing wrong with that. If somebody was interested in understanding the distinction there more deeply, then they would probably find themselves inventing a more refined terminology for it. Rather than say, "Is this really beautiful?" they would say, "There is a thing that we want. It is this. I'm going to explain it, and the cave painting has heaps of that, and there's this other thing which we want in a different context, which the people who did the cave painting also wanted but weren't very good at achieving." And if somebody was spending their life on teasing out that distinction very finely, then they'd probably invent a more fine terminology.

1:11:15 **Sadia Naeem** All right, Jesse? Jesse, you need to unmute yourself please.

1:11:30 **Jesse** There we go. Hey David, I have a question that might be somewhat personal but have a lot of implications in a lot of people's lives. I know Lulie has talked about this, it revolves around just romantic relationships, personal relationships, and the whole dichotomy of genes versus memes. We need society to procreate now, we don't live an infinite life, we know immortality is possible in some sense, but I guess there is a sense of, like, we want to create the best memes that we can, we want to create the best explanations that we have in our lives. How do you think about that in terms of children and education, whether or not to have a family or be in a relationship or just work on things like constructor theory and AGI and life extension or biotech? Or just really curious to see how you think about all those ideas.

1:12:40 **David Deutsch** I don't think it's a good idea to try and save the world in the sense of subordinating one's own values to what one thinks the world's values are. So maybe the world needs a larger population. My guess is that it does, in other words that would be a good thing, that the world as a whole would thrive better if it had more people in it. And other people of course think that the world would thrive better if it had fewer people in it. I think in both cases, it's a bad idea to subordinate one's own life to that objective. I don't think it's even, for example, a good idea in my own life to try to publicize my own ideas. I do it to some extent, but I don't subordinate it to the fun of actually trying to solve problems. Some of the problems are only of interest to me, some are [of] interest to me and like half a dozen other people in the world, and some are of interest to more people. But the way I would choose what to do is: try to meet my own values, and to the extent that my own values include having preferences about how the world is, then the meeting my values would include trying to make the world better. But trying to make the world better as an overarching principle for how to make personal decisions I think is a mistake. I don't know if that's your question.

1:14:44 **Jesse** Yeah, I guess that answers a little bit of it. And then it's just like being young, a big part of culture in general in society is just finding a significant other or a partner and there's the whole debate against polyamory and or to have a committed monogamous relationship, and that drives a lot of culture.

1:15:10 **David Deutsch** Yeah well, different people find answers in different ways, and they have extremely different problem situations.

1:15:19 **Jesse** From the context of *The Beginning of Infinity*, what was actually useful? It was useful to make more people, and to do that, people had families to do that in a kind of divide-and-conquer kind of sense whether they knew it or not, right? People, when they team up, they're more than the sum of their parts.

1:15:43 **David Deutsch** Yes. There are many ways of teaming up, and each of them has better and worse ways of doing it. So you form a society, you form friends, you form families, and all of those can involve mistakes in how to do it. We've got here by people making progress with that, but for most of history, they didn't make progress.

1:16:20 **Bruce Nielson** I was just going to say we are coming up on the hour, and I wanted to be cognizant of David's time. We still have it looks like quite a few questions, but how much time do you feel you've got left here, David? We'll kind of roll with that.

1:16:40 **David Deutsch** I'm willing to go on for a while, but I need to have another cup of tea. So if I could make myself a cup of tea, come back, then I would be willing to answer a few more questions.

1:16:54 **Bruce Nielson** Let's do that. I think that's a great idea.

1:16:57 **Unknown** Okay. A tea break, as it were. I'm going to go get tea as well.

1:17:06 **Sadia Naeem** I just had my tea, so maybe I'll have another one.

1:17:12 **Jesse** All right. Well, in the meantime, for all the non-tea drinkers, I guess we could just kind of shoot the shit.

1:17:20 **Bruce Nielson** Yeah, absolutely. So we're trying to go in order of raised hands here, by the way, so Sadia is the one who's the official moderator and she's got controls.

1:17:34 **Sadia Naeem** I'm getting tired. By the way, you're welcome to take over. I would much rather just chill and relax and listen.

1:17:41 **Bruce Nielson** Sadia, the problem is that you're the only one who can do it.

1:17:45 **Sadia Naeem** Yeah. Can you not see the hand raises, I guess?

1:17:49 **Bruce Nielson** I can see the hand raises. I can call on people. You'll have to control. Maybe they can unmute themselves.

1:17:55 **Sadia Naeem** Yeah, I could do that. I could take care of unmuting if somebody forgets.

1:17:59 **Bruce Nielson** Okay, I'll do that. I will call on people.

1:18:01 **Sadia Naeem** For some reason, I've always found that tiring.

1:18:05 **Bruce Nielson** I'll start calling on people if you'll take care of the muting. They can probably unmute themselves, but if they don't, then you're going to have to step in because I can't do it.

1:18:13 **Sadia Naeem** All right.

1:18:23 **Unknown** Holy hell, there's a lot of people in this room.

1:18:25 **Bruce Nielson** Yeah, there is. We've got fifty. I think we were at 59 at the top.

1:18:30 **Unknown** Wow. Can everyone unmute themselves if they want to?

1:18:37 **Bruce Nielson** We would prefer that people don't. Several people had background sound and we're just trying to make things easy, but at this point, while he's gone for a second, if anybody wants to shoot the breeze, that's fine. If you could just unmute yourself and re-mute yourself after you ask a question.

1:18:57 **Jesse** You are still a person who doesn't swear, I could tell.

1:19:02 **Cameo Duran** Some lessons get burned in really deep.

1:19:07 **Sadia Naeem** Actually, I was just wondering, Clovis, I know you gave a comment on that question about morality. I guess, in a sense, I sometimes find that [unsatisfactory] in physics, in a way we do talk about metaphysics, right? And when it comes to ethics, we're talking about metaethics. Kind of like when we're thinking, okay, is it instrumentalism versus maybe some form of realism? Because the questions that pop up are very much dependent on our metaphysics. And that's why I was kind of wondering that, even with morality, we may not address the question that what is the background metaethical, some sort of metaethics that's going on at the back of our mind. We could always just keep asking questions, but I think at some point there is some value in addressing that. Any ideas, any thoughts on that?

1:20:08 **Audience Questioner** I think one point to bring up is the concept of truthmakers. So traditionally, if you have a correspondence theory of truth, you think that there are true facts in reality and things correspond to them. Our theories can correspond to them. In theory, we're always fallible, of course. So the [question] is: Are there true facts in reality about values and morality?

1:20:36 **Sadia Naeem** I recently came across something interesting that Popper had said, according to him, "The values originate with life, just like with problems, as the problems arise." When he said life, he said that he means even before consciousness existed. "So all problems originated with life," I think he said, "and then values originated with problems," which I thought was interesting.

1:21:09 **Audience Questioner** Yeah. And also the question is not just about whether values exist, but whether they can be objective or not. So someone can have a value, but you can say, "That's just your opinion. That's just subjective."

1:21:22 **Sadia Naeem** No, and that is of my interest. Otherwise, yeah, we can come up with whatever, you know.

1:21:28 **Clovis** I agree with it as well. I'm trying to express myself in English and understand always. I'm sorry, this is tough. But right now, I am at war with fans of Sam Harris on this topic, because it's really something that I think divides us. David is back, I'm gonna give it back to him.

1:21:49 **Sadia Naeem** All right. Tracy, you want to go next? Tracy?

1:21:57 **Tracy** Sure. Hi. So I'm hoping this is just more of a fun, light-hearted question maybe, but on Thursday I woke up, I had a dream that I had gotten the opportunity to meet you, David. And the very next day, I find out that suddenly there's this opportunity to meet you at the Zoom meeting. Exciting for me. And kind of strange. So maybe the fun part, could you maybe speak to the human brain regarding its potential for quantum prediction maybe, or just the idea of quantum prediction in general?

1:22:41 **David Deutsch** So I'm not entirely sure what you mean by quantum prediction, but predicting the growth of knowledge is inherently impossible. And there's no reason to think that quantum effects might be implicated in the human brain. And the idea that quantum theory has kind of mystical, that it justifies various traditional mystical ideas, always comes from mistakes about quantum theory. It doesn't come, the real world doesn't implement those. So I think there wasn't a connection in that, I would guess that there wasn't a connection in that respect. Maybe that's a boring reply, but my guess is that's the truth of it.

1:23:50 **Bruce Nielson** All right. Mizrob. I don't know if I pronounced that right.

1:23:55 **Mizrob** You can hear me? Hi, everyone. Nice to meet you, David. So I just wanted to ask about replication crisis, especially in psychology and in general too, like in life sciences. So around 2010, people started to realize that there are a lot of studies that can be replicated. And so people started to implement many standards of like data sharing and open code and stuff like that. And there was also emphasis on importance of replication studies, like studies that repeat the experiment as closely as possible to the original study. So there is a sense that if a study is replicated, then it must be true. And less emphasis on mechanism—by mechanism, I mean [an] explanatory theory. They establish a link and by experiment, then afterwards give an explanation how this process might happen in the mind. But they prioritize replication, seems to miss the point that we can replicate, say, Newtonian laws [infinitely] many times, but they're not [an] actual explanation of how the world works around us. I just wanted to know how you see this, what you can say about methodology of psychological studies.

1:25:22 **David Deutsch** Yeah, I entirely agree. I think the replication crisis in psychology and related fields, as you have just said, I think it's the wrong way to think about it. [The] replication crisis is a small facet of what goes wrong when you apply scientism to psychology and anything that involves knowledge, anything that involves human knowledge. If you try to study it as if it were physics, you will be doing scientism, you will get it wrong. And the fact that it's not replicated is almost a positive feature of a theory, because it's at least saying that the explanatory part of the psychological theory, which was kind of unstated and taken for granted and implicit and denied and so on, [that] that thing existed, that there was an explanation there. And that's why the explanation can be falsified by an experiment.

If something can be replicated in psychology, then it's not really psychology. For example, people do wonderful work creating optical illusions and explaining why they work. And they work in psychology departments, many of these people, but that's not psychology. That is a study of the human visual system and how the information is processed, but that information is not being processed by a creative process. There are other kinds of things that stem from that, that you might ask. Then after the built-in interpretations of sensory data, there is further interpretation [that] happens, which can be creative and which also affects how we perceive things. And you can form theories about those, but those theories have to be explanatory, and there has to be a model of those. There, I would say that replicating them on a computer might be a useful thing to do with those explanatory theories. "If you can't program it, you haven't understood it," might be relevant there.

As I think you hint, I think the real trouble with psychology and related fields is scientism and a lack of, and even a denigration and deliberate avoidance of, explanatory theories. This was explicit in the case of behaviorism, but behaviorism has kind of been rejected. But the aspect of behaviorism that says, "One should not have explanatory theories, but rather one should have massive data which is replicated," that is still there and that's what really needs to be reformed.

1:29:21 **Bruce Nielson** Thank you. Dennis?

1:29:24 **Sadia Naeem** Hold on, I'm sorry, Bruce. Actually, there was somebody else ahead who [had] dropped out. Could I just call him in again? He sent me a message. Go ahead, John. Sorry about that.

1:29:37 **John** [Hi] David. Thanks for doing this. You had mentioned earlier, I'm speaking from Jerusalem, Israel. You had mentioned earlier the Popper lecture and later paper on the nonexistence of scientific method. I just thought you might get a kick out of this volume that I found literally lying next to a dumpster from 1958, which is apparently the first Popper piece of writing that was translated into Hebrew. I know you're from Haifa, so I thought you might get a kick out of that.

Anyway, my question is, in your first chapter in your book, in your theory on explanation, I've always wondered, I always got the feeling as you step through the phases leading up to the breakthrough method that we have today, which is, of course, one step in the long chain. I've always wondered how you see the relationship between that theory and Popper's. I [wouldn't] normally bring this up, but I know this is a Popper-oriented group. So I was just wondering if you saw that theory as a corrective, as completely 100 percent compatible with, and just another way of looking at it, or how do you see it relating to Popper's theory of explanation? Thanks.

1:31:01 **David Deutsch** So I privately and personally think that it is Popper's theory. I'm not a historian of science, and I'm not really interested in who had what idea, but I see, for example, the first chapter of *The Beginning of Infinity* is just a small explanatory footnote to Popper's epistemology. And if somebody comes along and says, "No, it's not, Popper thought something completely different," I don't care. I'm only interested in what the truth is. At the other extreme, if someone comes along and says, "That's exactly what Popper said, and even your footnote is in a footnote of Popper on page 483." Well, again, I don't care. I am trying to understand the world, and I'm interested in what's true. And attributing it to Popper is merely a matter of kind of academic courtesy. So I think that Popper had an entirely explanatory conception of science. I can't prove that from his writings. And I know that, for example, David Miller thinks that's not entirely true. Again, sorry if it sounds dismissive to keep saying I don't care, but it's not what I'm interested in.

1:32:47 **Bruce Nielson** Thanks. Thank you. Dennis.

1:32:51 **Dennis** Hey guys, can you hear me? Great. Hey, David, it's Dennis. Earlier you mentioned, in response to Ella, Ella was asking about self-replicating ideas in the mind. And your response was, if I understood you correctly, it wouldn't really be efficient in terms of memory, because instead one could have a quantity field of sorts on ideas that would encode how many instances of an idea exist. And then, that way, one could save a lot of memory. But I want to take a moment to defend the theory, if I may. As it happens, Ella has thought of the same thing when we first started discussing this theory.

Now, I suppose the quantity field would be denotationally equivalent to having replicators on the surface. But the structure of the implementation would be wholly different. And I think one would lose a lot of explanatory power by removing replicators, because one would need to come up with separate explanations for everything that the replicator-based explanation currently [explains]. For example, memories, how people evolved with some ideas survive in the mind, not others. So I'm not sure just because a programmer would prefer to use quantities instead of replicators, that means that biological evolution would have 'chosen,' I say 'chosen' in scare quotes, to do so as well. Most of the criticism of this Neo-Darwinian theory of the mind, if you want to call it that, that I've heard so far, is along those lines that we don't need replicators and that we could replace them with something else. And if I understood you correctly, your criticism is along the same lines, but the epistemological problem that I see with that is we could say that for any theory, right? I mean, even hard to vary ones we could think of ways to replace key components of them, even if usually that means that they become easier to vary as a result. And I think that's what happens when we drop replicators.

The problem reminds me a little bit of the fossil thing, which I believe you've brought up before in defense of the multiverse. So people might claim we don't need to claim that dinosaurs really existed to explain fossils, even though that is already a hard to vary explanation. We could simply come up with other ways fossils may have come about that don't involve the existence of dinosaurs. And then, denotationally, I suppose those theories are the same or at least similar because the output of the theories, the dinosaur fossils are the same or going a bit off the rails Instead of claiming that many dinosaurs existed, we could claim that there was a single dinosaur that had a quantity value that determined how many fossils left behind or something like that. Right? So I guess the problem is that this won't convince the advocates of the past existence of dinosaurs rightly, I think, because they would want to know why dinosaurs couldn't have existed, not why they need not have existed. So in a way I agree that dinosaurs need not have existed for the same reason that no theory need necessarily be true. And so that applies to self-replicating ideas of the mind as well. But what I'd really be interested in is a refutation, like an argument, right? Why replicators can't play a role in how the mind works. Can you think of such an argument?

1:36:17 **David Deutsch** No. I did say that I don't know how any of that works. And maybe you're right that maybe it's the fact that I learned programming a long time ago, and my formative programming years were in an era where memory was expensive and it was worth spending time, thinking of more efficient ways of storing the data. And now memory is extremely cheap and it's usually not worth doing that. And as you say, one of the things you gain when you have a redundant representation of something is you get much more flexibility in explanatory power. So having said that, I think your comparison with the dinosaur theory is a bit unfair. If your problem is that you want to make an artificial fossil, it would not be a good idea to start by making dinosaurs. You need to take the shortcut that's available and make an artificial fossil that way. And again, if you want to explain how the fossil got there, that would be a terrible way of approaching that problem. But if you want to make an artificial fossil, then going via dinosaurs is far too inefficient. But, since I don't know how it works, I can't really pontificate about how to do it. Let a thousand flowers bloom.

1:38:06 **Dennis** Got it. Okay. Thanks.

1:38:13 **Podge** Hi guys. How's it going? Thanks to Sadia and Bruce for putting this event on and for David for answering questions. So my question was about the explanation of how creativity works or just what creativity is. Critical rationalism in general seems to contradict certain commonly held assumptions, which are effectively just statements that people are mechanical. For example, operant conditioning, which is that learning and just alterations to human thought or the thought of people more generally and their behavior is best achieved using a framework of rewards and punishments. So that when dealing with problems in psychology, like maybe addiction and other, it seems to get a lot of uses within psychology and then in behavioral economics as well in the form of incentives and disincentives to do certain things. I think the original question I actually had, specifically about addiction and making choices, was sort of answered already when you were speaking about just creating the best moral theories and so on. But I was wondering if you could say something about incentives and disincentives, and how valuable the work done in behavioral economics is, and whether it's just fundamentally based on faulty assumptions and there is not much use to it, or it's just maybe contingently useful based on the cultural ideas at a given time or something like that.

1:40:09 **David Deutsch** Yeah, so I have to recognize that lots of things in the world do not involve creativity, and such things can be analyzed in terms that would be dehumanizing if applied to things that do involve creativity. And economics, for example, is a field where sort of the important issues are dominated by creativity, but not totally exhaustively described by creative processes. There are other processes as well. And if you're looking at an area of the economy where not much creativity [is being] used because people find the setup basically satisfactory, and what they want is a mechanical way of getting through to various things, then you can find an algorithm that sets the prices in those situations. You know, like, when there's a shortage of some raw material, then you can work out how, at least the first idea of how you can set the price, although someone else might think of a better idea, and already you haven't modeled that.

And similarly, if there are things that happen in the human mind, in the human brain, I should say, that aren't creative, like optical illusions and that kind of thing, and if they feed into the problem that you have, which is partly about creativity and partly not, then that might be helpful. I'm not going to say that isn't helpful, but I say that whenever creativity touches on something, it changes it profoundly, and it really becomes the most important thing to try to understand in regard to that field. Rewards and punishments are an abomination, really, in anything to do with humans, because they are trying to forcibly change a human situation, which had involved some creativity, to one that doesn't. And that is just bad.

I wonder, it's like these purported cures for gayness and so on, by giving gay people electric shocks, and if people want to be treated like that, they are making a mistake. I don't care if it 'works' or not. Works, in quotes. I'm wondering: How would you cure, if you thought that an S&M fetish was bad for you, and you had one, and you thought it was bad for you, what kind of conditioning would you expect to cure that? You know, being given electric shocks when being given electric shocks.

1:43:43 **Bruce Nielson** Sorry, please continue.

1:43:46 **David Deutsch** Sorry, I was just making an extended analogy that I thought was quite amusing, but it may not be interesting.

1:43:57 **Bruce Nielson** Thank you. Thank you. Karl.

1:44:00 **Karl** Hey David, thanks for doing this. It's been really fun. So I remember you saying in an interview that whether animals suffer or not is a philosophical question rather than a scientific one. And I definitely agree. So I'm just curious to hear if you found any convincing arguments for either side of that issue. And if you haven't, how do you think we morally should treat the issue of whether animals suffer or not?

1:44:30 **David Deutsch** Yes, I think not much is known about this. I think there are some tiny clues in various places. And I think that maybe the main thing is, since we know so little about this, I think there is room for a range of views that can all be considered reasonable, depending on where one is coming from. One can rule out, I think, the extremes, like thinking that, on the one hand, we should respect the wishes of trees is very close to being untenable philosophically because of what we know factually. And at the other extreme, I think that it is wrong to adopt a position of principled callousness and [try] to abolish, for example, all laws about animal cruelty and whatever, on the grounds that there's no evidence that anyone is suffering when there's animal cruelty. There is no evidence. But I think that is different from saying that there is a good reason for adopting that view. But in between those extremes, there's a huge range of positions that I think are reasonable.

1:46:26 **Karl** But would you say that this is a mild form of the precautionary principle, that in the absence of knowledge we should try not to...

1:46:35 **David Deutsch** No, well, I think it's more that what we should do in the face of ignorance. In the face of ignorance, the first thing is to be tolerant of multiple views, and the precautionary principle precisely isn't. So, I would say be tolerant of multiple views about this. You said about evidence, [there's] a tiny piece of evidence in regard to dogs. Dogs look like they have feelings more than similar other animals do. And we know that this is because they have been subjected to artificial selection for precisely the attribute of looking as though they have feelings. Now, I'm not sure that looking as though you have feelings can be done without having them. This is a very weak argument. I can easily think of ways [that] that might not be right. But beggars can't be choosers. I think we have touches of evidence that maybe some animals have some element of qualia. But if this counts as anecdotal evidence of something, there is strong anecdotal evidence the other way as well. If you look at animals like chimpanzees that look as though they have feelings, in other experiments, it's fairly clear that they do not have an idea of what's going on. That they're just behaving mechanically.

1:48:51 **Karl** But you tentatively reject the notion of philosophical zombie dogs, then, I guess?

1:48:56 **David Deutsch** Yes. Because that's one of these all-purpose explanations that could be used about anything. I can imagine a theory with a physical zombie Jupiter, where Jupiter doesn't exist, but only looks as though it does. So that's a whole class of explanations that have to be rejected on principle.

1:49:23 **Bruce Nielson** All right, thank you, Karl. Cameron.

1:49:30 **Cameron** Hi, David, can you hear me? My question is sort of around my trouble reconciling Popperianism, Deutschianism with behavioral genetics, namely that it seems to conflict with universal computation. I think you've noted that your position is that the mind is not a blank slate. So we have inborn genetic knowledge, but importantly, that can be overridden or overwritten. Examples such as fasting, celibacy, and skydiving, and suicide. But so my understanding of the behavioral genetics literature is that genes seem to predict many behaviors. I think a lot of people in that field may say explain, which I think you have issue with. And over the last fifty years, the main evidence of that is around identical twins versus fraternal twins, identical twins being more similar, siblings being more similar than adopted siblings, and adopted children being similar to their biological parents and not similar to their adopted parents. I think Robert Plomin describes genes influencing behavior as, describes what is rather than what can be, which I think aligns with one of your comments around the amount that genes influence our behavior is itself a product or function of culture. But I think your position is that genetic knowledge or genetic influences is probably easy to be overridden and probably happens early on. So I have trouble reconciling that with, I suppose, the fact of the adopted children being sort of systematically similar to their biological parents, their particular biological parents. And it seems to me that genetic influences do have a very large influence over what currently is. So yeah, if you just want to react to that.

1:51:55 **David Deutsch** Yes, I think that the experiments on twin studies and sibling studies and so on, correlations between behaviors of genetically similar and environmentally similar, none of those experiments addresses the issue. Put it in computer terms, where is the code located that is responsible for those similarities and differences? And where did that code come from? Given that, as you just mentioned, given that the degree of genetic influence on behavior is itself determined by culture, that alone means that you can't do a behavioral experiment to distinguish cultural from genetic behaviors. Sorry, you've got to be very careful in talking about these things. You can't do a behavioral experiment to distinguish between differences between different people's genetic or cultural knowledge. And so in regard to this issue, I would just reject the relevance of all those experiments.

I think there is a very strong argument, as you just said also, that genetic behaviors, again, the differences between genetic behaviors of different humans are relatively easy to override. I don't mean that one can override them oneself just by waking up one morning and deciding to. On the contrary, that might be very hard. But, for example, memes, either rational or anti-rational memes, can just not override but just replace genetic behaviors systematically, because they have evolved the knowledge of how to do so. And there are cultures where people are more or less careful about dying. It's not to say that someone in that culture or someone in a different culture could change that setting at will.

But on the other hand, I think it provides a very strong argument for saying that if that is a problem that one has, it is soluble. One can alter one's inborn tendencies in the same way that one can alter any other idea that one has that affects one's behavior. One can have a habit of writing with one's right hand, and then if one's right hand becomes paralyzed from some illness, one can learn to use the left hand. And one can't do that overnight, but one can do it, and one can do it arbitrarily well. And there are ways of doing it faster or slower, and there are always ways of improving those ways, and so on. Right, I think that the genetic explanations, while one can always form genetic explanations, I think they are, in regard to behaviors that are changeable, those explanations are dehumanizing and false.

1:56:23 **Bruce Nielson** Thank you. David, got another forty minutes since the last time you took a tea break. How are we doing? How are you doing?

1:56:37 **David Deutsch** Maybe we should draw things to a close soon. I don't know.

1:56:42 **Bruce Nielson** Okay, how about we do two more questions? Is that okay? Okay. All right, Bart.

1:56:54 **Bart** Hi, thank you, David. Thank you, Bruce and Sadia. Actually, tomorrow is my birthday, so I guess this must be one of the most original birthday presents to get to ask you a question. My question is the following. Is our society open enough for us to, at one point, refute justificationism in favor of critical rationalism, collectively enough? And what do we have to imagine as kind of acceleration effects on the growth of knowledge when that happens?

1:57:27 **David Deutsch** Well, happy birthday. If we're to be rigorous doctrinaire Popperians, that's a joke, then we shouldn't ask, "Is society rational enough to accept critical rationalism?" We should ask, "Is society capable of making progress?" Because we don't know that critical rationalism is true. We don't know that what we think of as critical rationalism really is critical rationalism, as perhaps there's a better view of it that is different from our view, and so on. So the question should be, "Is society capable of making progress?" And I think it obviously is, it is making enormous progress. The things that worry us about when we notice that some things are going backwards, it's natural and good that we should focus a bit on those, rather than go on about how well things are going. We should be focused on problems and things going backwards, in some respects, is a problem and deserves having creativity devoted to it. But overall, the big picture is that there's enormous progress being made at a rate that's unprecedented in history. So yes, I think there is such progress. I think that society can, although it may not, people on the whole may make the wrong decisions and everything may go wrong. But it is possible for things to go right. And I think at present, they still are going right on the whole. So I'm optimistic.

1:59:37 **Bruce Nielson** All right, thank you. And then final question, Aaron.

1:59:41 **Aaron** Thanks so much. I read an interview where you described being messy and untidy in your home, but being very rigorously organized on your laptop. And I couldn't follow what the distinction was. Why is [it] orderly in one domain and not in the other?

2:00:10 **David Deutsch** I think I was going through a phase of experimenting with the MacOS and noticing how pre-thought out and sophisticated the model was. [Of course,] it is nothing compared with today's. And also it's not just the Mac nowadays that has those things. I think nowadays I'm pretty sloppy in my management of my computer as well. So I'm sloppy in all ways. And what's more, I think if I can make a personal self-criticism, I think I'm too sloppy in most ways. There's some kind of irrationality there. But being very sloppy compared with the norm on a computer or in one's mind or in one's home or in one's office and all those things is useful for most people, most of the time, for the reason that I said in that interview long ago. The reason is that imposing a structure is a theory. And it includes inexplicit theories. And if one takes a view on that that's too rigid, then one is putting a strain on the possible new ways of thinking about that that one can explore.

2:02:04 **Bruce Nielson** All right. Thank you. David Deutsch, thank you very much for joining us. I know I really enjoyed this. I can tell this has just been a fun chat for most of us. So thank you for showing up for the Karl Popper meet and greet.

2:02:20 **Sadia Naeem** Thank you very much. Thank you, David. Just wondering, by the way, did you have anything to do with writing the script for Pickle Rick? For Rick and Morty by any chance?

2:02:31 **David Deutsch** No. I wish I had.

2:02:33 **Sadia Naeem** Someday, not today, but someday I wouldn't mind asking you what if Pickle Rick found himself on Earth which suddenly transformed into a planet made of cheese, do you think you'd be able to survive?

2:02:47 **Bruce Nielson** The consistency of cheese. The consistency of cheese.

2:02:51 **Sadia Naeem** Some other time. Just wanted to leave you with that.

2:02:54 **David Deutsch** Maybe if you do this again next year, you can invite the author of that episode because whoever the author or authors were, they got that amazingly right. It's like a manifesto for human creativity.

2:03:13 **Various** All right. Thank you, everybody. Bye. Thank you. Thank you. Thank you. Thank you, Bruce and Sadia for running this too. Yes. Thank you. Thanks for your participation. Thank you. Thanks, guys.

SAM KUYPERS:

MUSING ABOUT STATEMENTS, PROPOSITIONS, AND TRUTH (WITH LIBERTY FITZ-CLARIDGE)

Hosts: Sam Kuypers & Liberty Fitz-Claridge

About the hosts:
Liberty holds Master Degrees in both Philosophy and Applied Linguistics and English Language teaching, as well as a Bachelors in The Philosophy of Science. As well as teaching, Liberty runs the Popper-Deutsch Reading Group which can be found at www.meetup.com/popper-deutsch-reading-group/.

Sam is a quantum physicist interested in foundational questions. In his research, published in reputable scientific journals, he explores the quantum theory of time, the many-worlds interpretation, and locality. His work connects these areas under a broader commitment to realism—the view that a real world exists and that science is a quest for good explanations describing this real world.

Sam's Substack: https://substack.com/@samuelkuypers
Liberty's X: https://x.com/l1berty

Organization: Oxford Karl Popper Society

About the organization: an Oxford-based student society with the aim to facilitate discussions about the topics that concerned our society's namesake, such as philosophy, science, politics, and education.

Organization Facebook page: https://www.facebook.com/groups/karlpoppersociety/

EPISODE DETAILS

Date	March 1, 2021
Hosts	Sam Kuypers & Liberty Fitz-Claridge
Source	YouTube
Episode	In conversation with David Deutsch: musing about statements, propositions, and truth
Description	A conversation with David Deutsch about his theory of statements, propositions, and truth.
Link	https://www.youtube.com/watch?v=DZ-opI-jghs
Ideas	• David's correspondence theory of truth refers to the correspondence between an abstract thing and a real thing, both completely unambiguous and completely objective. And our statements about them are always imperfect attempts to express them well enough to solve whatever problem we're addressing. • We're limited in what we can model, but to think of knowledge as just a model is not true. We can talk about things that we can't model. We can understand things that we can't model. • In some cases, you can say that one theory is unambiguously better than another theory, because the set of true implications of one of them includes the set of true implications of the other and vice versa—the set of false implications is contained in the set of false implications. But that's not always the case. Sometimes, the set of true and false propositions just overlap. But if you think not in terms of truth, but in terms of knowledge, then when we have eliminated some errors, and hopefully not introduced other errors, we have unambiguously made progress, regardless of what the true implications are of the relevant theories.
Topics	abstractions • choices • correspondence theory of truth • Danny Frederick • decidability • explanations • explanatory argument • fallibilism • metalanguage • models • Popper • proof • propositions • realism • statements • Tarski • uncertainty

Episode title and description reproduced verbatim from the original source.

Transcript

0:14 **Sam Kuypers** Good afternoon, everyone. Thanks very much for joining us. Today we have David Deutsch, author of *The Beginning of Infinity* and *The Fabric of Reality* and physicist at Oxford, and David will talk about truth and propositions. And we have a slightly different format today, this will be a conversation. So David will give a brief couple of remarks on the topic, and then we will have a discussion with David. First, Liberty and I will join into the discussion, and later on we'll open it up for everyone. And with that being said, I give the floor to David. Thanks for joining us at the Popper Society.

1:00 **David Deutsch** Hi. Well, thanks for having me. It's statements, propositions, and truth that I'm going to muse about. And the context in which I was musing is, first of all, Tarski's theory of truth as adopted by Popper, which is called 'correspondence theory of truth.' And the idea is that a statement is true if it corresponds to the facts.

[writes "This is a dog" beside a digital image of a dog]

Now, there's a statement and there is a fact, and that statement is true if and only if that really is a dog. And that's the correspondence theory of truth. And I thought it was satisfactory, and I believe Popper's treatment of it. But I started musing further about this in conversations with Lulie Tanett and other people and also because of the talk that was given here by Danny Frederick about a slightly different issue, whether truth can be our epistemic aim. Not quite sure what he meant, but in any case, I realized that my own conception of truth was somewhat flawed and I came up with some ideas to fix it, which I could have called "a simple theory that resolves all your misgivings about truth." Of course, it might not resolve all your misgivings, it might not be true, or it might be true but not original to me, in which case it's most likely to be found somewhere in Popper.

3:16 **David Deutsch** Anyway, so there's a problem. What is the problem? Well, for a start, that's not a dog. That differs from a dog in a number of really vital ways. For a start, it's a cartoon. No dog actually looks like that. And secondly, when I say, "This is a dog," how do you know that I'm referring to that dog and not to some elephant that's in the room? So it might be referring to the elephant in the room and then it would be false. And so I could make it more precise by saying, "On this page, this is a dog." That's more precise. But you see the point—I could go on adding qualifiers and explanations and so on ad infinitum and it still wouldn't be completely unambiguous. Yet reality is completely unambiguous. So how can an ambiguous thing correspond to an ambiguous thing? And so that's one thing that's the problem.

Something that's perfect and objective but not directly perceptible, that's the dog, how can that correspond to something that's imperfect, parochial, and perceptible, which is any kind of statement about a dog? Anything physical is like an idea in our brain or a statement in words, which can never be perfect [and] so it can never be perfectly true. Nor can it be perfectly precise and perfectly unambiguous. By the way, [there's] also a metalanguage involved in this theory of truth, which says something like, "This is true if this is a dog." So that's a statement in the metalanguage. So there's a sort of triangle here of a statement and some reality and a meta statement. That's Tarski's theory.

6:00 **David Deutsch** And another reason why, apart from being ambiguous and so on—we are fallibilists. We expect to improve our ideas. And so the ideas in our brains and our statements of them can't be perfect if they can be improved. So here's what I thought might resolve this. In addition to statements and some reality, by the way, the reality could be mathematical reality and so on. But I'm using physical reality to simplify things. We could be talking about prime numbers and exactly the same issue would arise. So there are some things here, there are some other things which are perfect and pristine and beyond our reach but which we can nevertheless talk about and form theories about. And that is the class of abstractions. Something like what Popper called World 3. But I just prefer to say that the class of all abstractions, which [includes] things like numbers, like so it's up here, there's a number five. That's an abstraction. And again, I can't draw the number five. I can only draw some marks on my iPad screen. That thing there is just the marks on a screen. And it's not a five, because, for example. That is not a five. It's a fifty, even though it looks exactly like this five. So again, anything I might draw or say is ambiguous and partly wrong and all that stuff. But the thing I'm referring to here, the actual number five, that's a perfectly definite thing. And there are also up here propositions, the propositions like, "Five is a prime," is a proposition. And except what I just said is a statement. I can't say propositions. It's impossible.

So there are propositions up here, which I'm representing. So there's a 'P,' and a 'not P,' and 'not P' and 'not P.' These are all propositions. And propositions have the property that they can be true or false and nothing else. Absolutely nothing else. [There's the] excluded middle. And they're perfectly precise. They're perfectly unambiguous. Here are some propositions. And this one is actually true. I mean, the proposition, this one proposition I'm referring to is actually true. But unfortunately, that's because it doesn't assert anything about anything. And also this 'P' and 'not P,' strictly speaking, those aren't propositions, either. They stand for propositions. They are propositional variables. So this thing is true because it's true regardless of what 'P' stands for. But just this one is a propositional variable and it stands for something that's either true or false, but it itself just represents that. it's a variable.

9:42 **David Deutsch** By the way, it's quite usual in talking about the world and about ideas that we're super used to things that stand for other things, and calling the things that stand for other things, the things. And sometimes we have to be careful and make sure that the map isn't confused with the territory, as they say. Usually one can disambiguate sufficiently well, but that's just another level of ambiguity that cannot be perfectly resolved. So if I try to write down here an actual proposition, [the] traditional ones are things like, "All men are mortal," and it suffers from the usual ambiguity, like whether this refers to people who were alive at the time of Aristotle or something, [it's] not clear whether 'men' includes women or whether men includes men who've already died or men who have yet to be born, and so on. But nevertheless, there is an idealization we can think of, which is qualified by an infinite number of qualifiers, as it were, enough to make it which we could never actually do in real life. But this kind of represents a real proposition somewhere in there. And this proposition could be true of the world with people in it and so on. And they're either all mortal or not all mortal. And then that's true if they are all mortal.

So the more you try to define this, the more vague it gets, but like with the five and so on, there is a real thing, a real abstract thing. And here's another real thing, the world, and they could correspond with each other. And my idea is that the correspondence theory of truth refers to this correspondence—the correspondence between an abstract thing and a real thing, both completely unambiguous and completely objective. And our statements about them are always attempts to express them well enough to solve whatever problem we're addressing.

So I can say that there are the abstractions, and then there's the reality, then there's a correspondence, and then there's a meta-theory saying, "True if they correspond." There's [a] meta-theory and it refers to those three things and then their statements. These are the nasty, dirty things that we can actually say, and above this line is pristine truth and exactness. And the miracle is that we can actually gain knowledge about these things. We can say things about them, never perfectly. So we have here instead of just statements corresponding to reality, which is what I thought Tarski's theory of truth ought to say, it's a triangle. It's this, this, and this is a triangle. And, if you like, this meta statement, I have to somehow bring it out of the page to form a tetrahedron. So it's either a straight line converted to a triangle or a triangle converted to a tetrahedron. And that is my theory of truth, which it is. And now you can shoot it down.

14:54 **Sam Kuypers** Great. Thank you so much for the introductory statement. So the first question I have is: How can it be that these infinitely long propositions can be approximated by finite statements?

15:17 **David Deutsch** Yes. I don't think there's any guarantee that they can. So I think the situation is worse than what you say, because saying that the abstract perfect thing is like the imperfect thing with an infinite number of qualifiers, I don't think there's any guarantee that even an infinite number of qualifiers would—even if that were possible, which it isn't—that even that would do. So how can we expect our statements to correspond to this pristine truth? Well, one thing is there is no guarantee they might not be able to. I argue in my book that there shouldn't be any limit to how well we can do this, even though we can never do it perfectly. And the thing to remember as always with Popper is, when Popper says, at least I think, when Popper says, "We're striving for the truth," he always means, "We're striving to eliminate error." Of course, we can't be sure of that, either, but when we use truth as a value that we are aspiring to, we're not aspiring to be able to utter true statements. We are aspiring to eliminate false statements from the things that we say, which always leaves more false statements. And we can't be sure that we haven't eliminated a true statement. So there is no guarantee, and there's no reason, not just no guarantee, there's no reason in the structure of all this, if I'm right, that implies that there should be statements that represent an abstraction perfectly, even infinitely long statements. So I don't know if that's good enough, maybe it isn't.

17:45 **Sam Kuypers** But so you think that we can guess at the correspondence between the statements and the abstractions, and that is how we kind of aim to find out about reality?

17:55 **David Deutsch** Yes, yes. We guess that our statement is some kind of indicator of the abstraction. Like if I say two plus two is four, that is in the light of a whole slew of theories that connect statements in the English language to integers. And then there's [a] theory of integers, which we can't prove true, either. It might be inconsistent for all we know. But somewhere in the set of all abstractions, there are consistent things. And we have a theory that the set of all integers is a consistent thing. And that what we say about it [corresponds] to true propositions about that thing. And in all those ways, we're bound to be inaccurate and ambiguous. But nevertheless, we might have some genuine knowledge about it.

18:54 **Sam Kuypers** Right. And I see people are raising their hand in the chat. I also think that Liberty has a question. So Liberty, go ahead.

19:05 **Liberty** This might be the same as Sam's question, but is there anything specific in Danny Frederick's kind of skeptical view that you think this resolves?

19:21 **David Deutsch** Well, I must say, I didn't entirely understand his skeptical view. It was directed against something slightly different from what I'm talking about, namely the question of whether we can regard science or just thinking in general as a quest for truth. And he was saying we can't. Now, I don't see that's a very important question. I'd rather say what we are doing rather than say whether it's legitimate to call that a quest for the truth. His was a much longer statement than my musing here. I think he also touched on the problem that I'm talking about here, namely that we can never capture a proposition, true or false, and that therefore we can never capture a truth. But I don't think that's important. As long as we know that we're not trying to capture a truth. We're trying to eliminate errors. And I think someone asked him that, and he answered, and again, I didn't quite understand the answer.

But anyway, I think this is a simpler issue. My problem is: If truth is correspondence with the facts, how can a statement which is incapable of being true or false correspond to a fact which either is there or not? In real life, there is an attribute, "being a dog," which any given thing has or doesn't have. We can't specify that unambiguously, but we can specify it unambiguously enough to meet whatever problem we're trying to solve. And this sort of higher-level thing of whether that counts as pursuing truth, well, I don't really mind. I don't really mind either way.

21:55 **Sam Kuypers** Right, so this is not trying to address the kind of skepticism that Danny was advocating.

22:01 **David Deutsch** I don't think so.

21:55 **Sam Kuypers** You already explained it quite well, so I understood your construction from that conversation with Danny, and I kind of saw [it] as a defense of realism and of the aim of science being the search for truth. So I think I had a misconception about what your construction was trying to address. And I think because of that misconception, I also misunderstood the problem you were trying to solve. So could you re-explain very briefly what the problem is that you're trying to address with your construction?

22:43 **David Deutsch** So, here we have some reality that's pristine, that's sort of perfect. And we have some abstractions which are pristine and that's perfect. And therefore it's meaningful to say that there is or isn't a perfect correspondence between them. When the correspondence is perfect, we can say that this abstract proposition is true. And then, there are these statements which are riddled with error and ambiguity and so on. And the problem was: How can one of these [Statements] possibly correspond to any of that [Reality]? The answer is: It can't and it doesn't. It's this [Abstractions] which corresponds to that [Reality]. And we are merely guessing that this [Abstractions] is what this [Reality] is and what the correspondence is. And we're just hoping that our guesses are better than our previous guesses. Our statements don't actually have to correspond to anything in the sense, in an exact sense, which would be required if we're defining truth that way.

24:23 **Sam Kuypers** Right. Would it be fair to say that you are defending realism?

24:28 **David Deutsch** Oh yes, but I think Danny would say that he's also defending realism. But we might think he isn't.

24:35 **Sam Kuypers** Yes.

24:37 **David Deutsch** But I really am.

24:40 **Sam Kuypers** We fallibly guess that you really are. I see there's some more people who have questions, specifically Lulie has her hand raised. Lulie, go ahead.

24:52 **Lulie Tanett** Yeah. So usually we talk about abstract things which are like philosophy and maths and then practical, down-to-earth things like this water bottle or whatever. And you're saying that all statements correspond to abstractions. Is this like a technical sense, or is there a meaningful difference between abstract statements and these kind of practical statements?

25:22 **David Deutsch** I think whenever we say anything about the world, we're doing it via an abstraction. We don't usually say so or even think of it in that sense, but that's because of our just being totally accustomed to having things which stand for things which stand for things which stand for things, and just speaking as if the first one was the same as the last one. But in reality, if you want to make sense of what theorizing means and what truth means and what having a theory about the physical world or about the abstract world means, then, according to my theory, you've got to say that all our statements are referring to abstractions which refer to the thing that we're talking about, rather than just referring to the thing we're talking about.

26:31 **Lulie Tanett** So there's no such thing as a purely abstract statement or a purely nonabstract statement.

26:37 **David Deutsch** I see what you mean, yes. We can't utter an abstraction. We can only utter physical objects like sound waves and so on. So we are purely physical, and our theories are always expressed purely physically. But we can theorize about, let's say, the world and about abstractions. And when we say, "This is a dog," and we say that in the context of some problem we have, then, because there is no such thing as a correspondence between these statements and real objects, like physical objects like dogs, but there is such a thing as a correspondence between an abstract proposition and a physical object like a dog. Therefore, we must always be talking about that abstract proposition whenever we say things that we're hoping are true or truer or whatever. If the concept of truth applies, we must be talking about propositions because it's only propositions that can be true or false.

28:01 **Sam Kuypers** Right, then I think I have another question, which is that, or unless Lulie wants to expand on her question, I hope I'm not interrupting.

28:13 **Lulie Tanett** No, go ahead.

28:15 **Sam Kuypers** Your idea of a statement, is that purely linguistic? Or because, for example, you were talking about the cartoon dog at the beginning of the presentation, and that also seems to be in some sense a statement. So it seems to be about not just sentences that we can construct, but also like the cartoon dog is a statement that corresponds to some abstraction.

28:47 **David Deutsch** Yes, I don't know. Okay, I hadn't thought of that. You could call this 'dog'. You could call this mark that I made here, you could call that a statement. But it would normally be called more like something like a model or a representation. Statements, models, representations, and utterances and sentences written down—the important thing from my point of view here is that those are all physical objects, and so they can't have the property of being true or false. But yes, this cartoon is, for present purposes, a statement about a dog. But in this context, I want you to take that as if it was a real dog, so that I can talk about whether that other statement there on the left refers to it or not, corresponds to it or not. But I can't, no matter how I try to talk about abstractions or reality, I can't directly represent them. I can only represent them as physical objects that aren't them.

30:11 **Sam Kuypers** I see. Okay, thanks. Then Charles, I see Charles has his hand raised. Go ahead.

30:19 **Charles Bedard** Yes. Okay, thank you. Hi. Thank you, David. Well, it's not a question. It's many ideas that I think would be quite cute to relate with other ideas of yours. I recently came about chapter five of *The Fabric of Reality*, in which you speak about the Turing principle in a very grandiose way. I was kind of aware of the idea that computer programs can be put into correspondence with simulations of physical systems. But the chapter ends with the idea that not only physical systems can be, or virtual reality rendering of physical [systems] can be given by computer programs, but also virtual reality renderings of abstract entities. So all that mathematicians think, and "All men are mortal" can be also, in a sense, thought of [as] virtual reality renderings of abstract entities. So the quest of science as search for truth in this context of trying to get a map between our virtual reality renderings of the reality out there can actually make sense, no? Because this is what we're kind of trying to do, to get our ideas to correspond to the physical world.

31:58 **David Deutsch** Yeah, okay. I was agreeing all the way up to the end.

32:03 **Charles Bedard** Sorry, I don't have any precise questions, but I just think that maybe it can give you the idea of commenting on this link between the Turing principle and this theory right here.

32:20 **David Deutsch** Well, there's no direct link in that sense, I think, because the set of all things that we can represent in virtual reality is basically the set of Turing computable functions. And the abstractions that we can speak of are a much larger set than that. We can talk about, say, the real numbers, without making a virtual reality rendering of all real numbers. So I thought you were going to say, "Therefore, our quest to understand the abstract world is exactly the same quest as our quest to understand the physical world, and that they are both done by making guesses about these abstractions, which are propositions." And that I would have agreed with entirely. But I don't think it's the case that we are just investigating the computable. We can perfectly well investigate the noncomputable.

33:32 **Charles Bedard** But don't we investigate the noncomputable through models of real numbers?

33:38 **David Deutsch** Yeah, we're poor, imperfect creatures. We also investigate it through much more crude things than just models. We investigate it through things like neurons and moving lips and so on. It's kind of a miracle, if this is maybe an answer to Sam's question at the beginning. It's maybe a kind of miracle that something so crude and error-prone as the part of the physical world that we control can have such tremendous reach, not only in the physical world but into abstractions. I don't think it's the case that we are limited. We're limited in what we can model, but to think of knowledge as just a model is not true. We can talk about things that we can't model. We can understand things that we can't model.

34:44 **Charles Bedard** But how do we understand them if we don't give [ourselves] some theory in some virtual reality rendering of the real numbers?

34:55 **David Deutsch** For example, most mathematicians think that P doesn't equal NP. And it could be true that P doesn't equal [NP]. Suppose it's true and suppose it's not decidable. Well, we can still perfectly well have a theory that P doesn't equal NP. It doesn't stop us at all, the fact is undecidable. We can even argue for and against that proposition because argument isn't proof. By the way, proof is useless without explanatory argument.

35:33 **Sam Kuypers** Yeah, I think that's a very nice point because obviously mathematicians do research into which functions are and aren't computable. So they have to be able to tell something, they have to be able to describe those kinds of abstractions in some way. Otherwise, they couldn't do their research, we couldn't be having this conversation about them right now.

35:55 **David Deutsch** Yes, yes. And before they can prove anything, they will have had a conjecture. You know, a mathematician could spend ten years proving that such-and-such a proposition is undecidable. But that's because the mathematician will have had arguments in mind that not only that it is undecidable, but that there's a way of proving it and so on. And if he fails to prove it, he may still think it's undecidable and I think 'P doesn't equal NP' is a good example of how easily that can be true. Everybody believes that, and it could be that nobody will ever prove it.

36:39 **Sam Kuypers** Yes, that's an excellent point. Yeah, Charles, I'm not sure if I'm interrupting your conversation.

36:46 **Charles Bedard** No, it's good. It's good. I'll think about it. Thank you. Great.

36:51 **David Deutsch** And by the way, another thing is just that the arithmetic of the integers is consistent. Again, everybody believes that too.

36:59 **Charles Bedard** But it's unprovable.

37:00 **David Deutsch** Yes.

37:02 **Sam Kuypers** I think there's various paradoxes that pop up here. Anyway, I see there's another question in the chat by Jake. I'm going to mispronounce [your] name. Sorry, Jake Orthwein. I hope I said that right. Go ahead.

37:22 **Jake** Thanks. So I have a question about this relationship between, I guess it could apply both to the relationship between our statements and the abstractions and to the relationship between the abstractions and the world. I'm more concerned about the relationship between the abstractions and the world. But I'm wondering about what it would mean to say [that] the abstraction corresponds to the world or even refers to the world, that certain pieces of the abstraction pick out certain things in the world and not others to refer to them, and what the nature of that relationship is. Independent of whether it's true, what does it mean to even refer for an abstraction?

38:01 **David Deutsch** Well, you're quite right that there are abstractions out there that don't claim anything. Let's say you could have the mathematical model of the Standard Model of particle physics. And let's suppose that that was true, it still wouldn't be an assertion. Whether it corresponds to anything depends on what it's claimed to correspond to. Because if somebody claimed [that] that mathematical model is a mathematical model of the weather on the planet Earth, then that would be false. So the claim is another abstraction. So the proposition would be something like: "The real physical world consists of fields and particles that obey these equations," and then some equations. And then that abstraction would have made a claim about the real world.

39:22 **Jake** But doesn't that threaten a kind of infinite recursion there, where it's never actually getting explained how it is that, say, the referent of particles picks out certain things in the world and not others. So it's not so much whether this description of particles is true of the world, it's: What does it mean to pick out some subset of the world and designate that 'particles' and therefore have this reference relationship between...

40:03 **David Deutsch** Part of the implication of this proposition that I'm imagining is that there is a physically real world. So that's one thing it would have to imply. And then it would say that "In this physical real world, there are things like electrons and so on, and they obey these equations." Now, it could be…that still doesn't say what distinguishes the physically real world from anything else. And it doesn't explain why there is one and only one of those. And there could be none of those, or there could be three of those, or whatever. But that's just a consequence of the fact that I can't imagine this perfectly exact proposition. It doesn't say that, in the world of abstractions, there isn't a proposition about the world which is capable of being true or false. If you think of the true one, then there is also just not that which is definitely a false one, though it's not an explanation. Because, as I'm always saying, the negation of an explanation is never an explanation. But the proposition which is an explanation could be true. It might not be the whole truth about the world. It's [just] talking about the Standard Model in the world. But that could be true. And if it claims that it is true, then it's a claim about the world, including a claim that there is one and only one world and physical world, I mean.

41:45 **Jake** I just have one more related question. So I guess it's not that there is a world that seems problematic to me, but the division of that into an ontology and kind of how you carve the joints of that ontology. And if you think about like a natural language statement like, "This is a dog." The reference relationship there could be accomplished just by appealing to the context. So if I communicate, "This is a dog," and you and I are in the same room and we're having the same kind of percept of there being a dog there, then the reference of "This" is just going to get worked out automatically by the context that we share. But when we're talking about the way reference relationships get determined between abstractions and the world, I have no idea how that would be accomplished. And then I also wonder if we stay in the same relationship to the abstractions as we stand to the world, which is to say, we don't have access to them. Why introduce the idea of the abstractions and not just talk about an error-prone, contextual problem-solving relationship to the world?

42:53 **David Deutsch** So in the sentence, "This is a dog." The word "This" is indexical. It gets its meaning from the context. The proposition that that refers to doesn't have a physical context. So it would have to explain what "This" is. It would have to say, "In the physical world, there is a planet with such-and-such characteristics and at such-and-such a time defined by these physical characteristics, there's a person sitting in a room with an animal which is in fact a dog," and so on. But it would say all that with perfect precision. So it couldn't say anything indexical. That's quite right. But I think the other part of your hesitation, I think I disagree with. You seem to be assuming that abstractions inherently must refer to other abstractions. They can't get out of the world of abstractions. But I think that is not so. I've just given an example of something where if it says, "There is such a thing as physical reality," then it is saying, I mean, [it] might be false, but it's saying something that isn't referring to just abstractions.

44:20 **Jake** No, it's not that I think that abstractions can't get outside of abstractions. I guess when you say, "There is a dog," my question is: What does it mean for that piece of the abstraction to correspond to some particular thing in physical reality? It's not whether there's physical reality at all. Whether physical reality comes divided into things like dogs or cups or whatever other kind of everyday scale.

44:49 **David Deutsch** If it doesn't, then the proposition would be false.

44:55 **Jake** But then you seem to be taking this reference relationship and bringing it back into this truth relationship. So is everything about whether a proposition refers to the world subsumed by whether it finally is true?

45:13 **David Deutsch** No, because it could be asserting something false about the world.

45:17 **Jake** But if to say that if 'dog' doesn't pick out anything in the world, it's not true and that's why it doesn't refer to the world.

45:31 **David Deutsch** It's not 'dog' that doesn't. So in my story, it started off by saying, "There is a real physical world, and in that world there was a Big Bang, in the Big Bang, there was a star, which we shall call the Sun, which…" and so on. And in each case, it gives enough context within the thing it is referring to define it uniquely and perfectly. Now, if at any point, that doesn't correspond to reality, then the proposition is false. It's still asserting something. It could say that "There is this planet that's loose without a star. And in that there's a person called David Deutsch who's referring to a..." Now, as soon as it has said that it is false, but it's still an assertion, it is still in my imagination a perfectly precise, meaningful assertion. It's an assertion about something which logically could be like that, but in fact isn't.

46:45 **Jake** So if I talk about "the Sun revolves around the Earth," the Sun has a referent in physical reality, but that statement about the Sun would be false. But I guess, independent of whether that statement is true or false, what does it mean for the Sun in that statement to refer to the thing in physical reality? So, what does it mean to pick out a thing in physical reality for the abstraction to refer to?

47:16 **David Deutsch** As I said, the trick I thought of, and there are probably many other tricks, is to first refer to the whole of physical reality. And then define unambiguously various bits of it until you zoom in on the one you want to talk about.

47:33 **Jake** So in this world of abstractions, something like a cup gets built up from the whole of physical reality all the way up?

47:41 **David Deutsch** Yeah, that's one way to do it. There might be a more efficient way.

47:44 **Jake** Thank you.

47:48 **Sam Kuypers** Okay, then I think Toby is next. Toby, go ahead.

47:58 Toby Yes. In your book, I think you [wrote] that "There's only a finite number of abstractions that could apply to the physical world, namely the computable abstractions and..."

48:13 **David Deutsch** [A] countable number. Yeah.

48:14 **Toby** The countable. So is it possible, if there is a finite number of abstractions that could apply to physical theories or unique physical theories, that we could rescue Popper's theory of truth-likeness so that we can get close to the truth because there's only a finite number of different physical theories we could discover?

48:41 **David Deutsch** Yeah, actually I think there's an infinite number of physical theories. It's just countably infinite rather than finite. Unless you are thinking that a finite region of the universe can only contain finitely many distinct states, which might be the case from the Bekenstein bound and so on. So, yes. But I think even then it would be pointless to resurrect Popper's notion of truthlikeness, because this would allow you to sort of say whether two theories are lexicographically close to each other in the dictionary of all possible statements, but that's not what Popper meant by "being close." He meant, "the set of all implications of the one is somehow close to the set of all implications of the other." I forget how it goes. It's the true implications of one other than tautologies. He was making it up in that kind of way. That would still be an infinite number. And I think it would still be infinitely ambiguous, it would depend on what your purpose was in comparing these two theories. One of them might be closer to the kind of truth you want to talk about. And the other one might be close to the kind of truth that someone else wants to talk about. For example, there's the kind of truth that leads to accurate predictions and then there's the kind of truth that leads to better future theories. So, why bother? I think that Popper in the end basically said, "Why bother?" as well.

50:45 **Toby** I've noticed with theories as they've gone through time, they changed the kind of invariant symmetry. So we had Galilean invariance and then we had Lorentz invariance. I was wondering, in that case, if we were seeking hard-to-varyness in theories. Perhaps there is a finite number of those different, unique variant mathematical structures which could apply to our universe. That was where I was kind of basing that idea from.

51:30 **David Deutsch** Well, so first of all, if there was just a finite number, but it was ten to the five hundred, we wouldn't be much better off. So if there was a finite number, and some people think that eventually there will be a single mathematical object, which is the only reasonable one to theorize corresponds with the physical world. And I don't think that would be the end of the story, either, because there'd always be the problem of: Why is that abstract object physically instantiated and not some other one? It couldn't itself contain the explanation of that.

52:23 **Toby** Yeah, I see what you mean there. Thank you.

52:30 **Sam Kuypers** Because Toby mentioned Popper's theory of truth-likeness: Do you think that, with this theory of correspondence to truth, that we can still talk about theories containing more truth over time? I know that this is slightly different from the problem you're trying to address. And I'm kind of drawing it back to this conversation we had with Danny Frederick on the same topic or on a related topic. And yeah, do you think that this solves anything?

53:04 **David Deutsch** First of all, my view is that in some cases, you can say that one theory is unambiguously better than another theory, because the set of true implications of one of them includes the set of true implications of the other and vice versa—the set of false implications is contained in the set of false implications. But that's not always the case. And in that case, the set of true and false propositions just overlap. But if you think of it not in terms of truth, but in terms of knowledge, then when we have eliminated some errors, and we hope not introduced other errors, then we have unambiguously made progress, regardless of what the true implications are of the relevant theories.

When we have successfully made a vaccine that cures the disease better than all previous medicines, then we have made progress. And we don't really need to measure how much truth it has. It might have been built on a theory of RNA, which is overturned next week, but where the overturning doesn't actually invalidate the explanations that led to the vaccine. So in that case, inventing the vaccine was genuine progress, is genuine growth of knowledge, even though it used a theory that was worse than the previous theory. Maybe that's a bad example, because these things have lots of different theories associated with them. But you see what I mean? I mean, science is about problem solving. So is life. And with problems, what we want to do is eliminate errors. If we can eliminate some errors, it doesn't matter how true the theory is.

55:45 **Sam Kuypers** Yes. I think that's a very nice reply. And I also have what is kind of a devil's advocate criticism of your theory about truth. We say that we can fallibly guess that there is this correspondence. For example, "This is a dog" corresponds to there being a dog on the page. And I think this is what Charles was alluding to. There are many cases where there is a lot of ambiguity and sometimes paradoxes that arise when we try to write down a statement. Like there's Berry's paradox, where you say something like, "The smallest positive integer not definable in under six letters, which is itself a description of that integer." And so you have this kind of self-referential paradox where the integer isn't well-defined. And initially, that seems to be a perfectly well-defined integer. I think when I first read that sentence, I go, "Yeah, that must be an integer that exists." But in fact, it doesn't. And there is a paradox that arises. How do we know that this isn't always the case, that there aren't many of these things plaguing our statements, that our statements are much, much too vague to ever reach out into the abstractions?

57:31 **David Deutsch** Well, again, we can't be sure. And if the arithmetic of the integers really is inconsistent, then we're talking nonsense most of the time. At least we're talking nonsense in the sense of logical implications of what we say. But not in the sense of problem solving. All our theories are false. That doesn't mean they don't contain knowledge.

58:21 **Sam Kuypers** Yes, I think in a way I was just re-asking the question I asked initially. And I just like the example of Berry's paradox.

58:31 **David Deutsch** Yes. Well, Berry. So that's one of many ambiguities that you can accidentally slip in a sentence from the metalanguage and mistake it for a sentence from the language because we use English for both. So it's an understandable mistake to make. So when you say, "The least integer not definable," you should be saying, "Definable within what language and what axioms," rather than just definable. Definable is a meaningless concept without saying what axioms you're defining it and what language in what you're defining it in. But that's as you say, "How do we know that there aren't infinitely many ambiguities like that which render meaningless everything we say?" Well, there could be. But we have a good explanation to the effect that we are, in fact, eliminating errors in our ideas, even if they're inconsistent. We're still eliminating errors from them.

59:45 **Sam Kuypers** Yes. And also, as I said, it was kind of a devil's advocate criticism because, of course, Berry's paradox is a very specific paradox, and we discovered it because other sentences aren't like Berry's paradox.

59:58 **David Deutsch** Right. Yes. But I took you to mean: that could arise and creep up on us. And there could be things that we don't know about that could also creep up on us.

1:00:11 **Sam Kuypers** Yeah, exactly. We have discovered a particular error in the case of Berry's paradox. And whenever those errors arise, we tend to notice them and correct them and then go on to the next thing, which is why we learn about paradoxes like Berry's paradox. So, yeah, there's more questions in the chat. I see Daniel, go ahead. Ask your question if you want. Otherwise, Teknu.

1:00:48 **Teknu** I have a question about mathematical truth, more technical question. You mentioned earlier the P equals NP problem, the possibility that this might be undecidable at some point. But leaving aside our abilities to settle those questions or not, consider the continuum hypothesis, which we know [from] the joint work of Gödel and Cohen that is undecidable from Zermelo-Fraenkel set theory with axiom of choice. I mean, that's a perfectly well-formed mathematical statement. Do you think that it expresses a proposition which is either true or false despite being undecidable? Because there are a lot of mathematical philosophers [who] think that this is just an indeterminate mathematical statement.

1:01:32 **David Deutsch** Yes. I disagree with those philosophers of mathematics. I think that a perfectly well-formed mathematical proposition is either true or false independently of whether it is decidable or not. 'Decidable' is in any case a matter of physics. So it seems to me ridiculous to base a theory of mathematical abstractions on what physics does or does not happen to be able to model.

1:02:04 **Teknu** Well, I'm not sure whether set theory is really about physics because it postulates a lot of infinities that go beyond anything that physics might study.

1:02:12 **David Deutsch** That doesn't matter. It's not the sets that have to be finite. It's the method of proof. So mathematics assumes that a proof is a finite sequence of propositions, each of which follows from the previous ones by rules of inference, which are also finite. There are finitely many of them, they're finitely long, and so on. And the proof is a finite one of those. And 'finite' here just means 'can be instantiated in a physical object.' So it's perfectly possible, logically possible, that physics is different from what the way we think it is, and that the rules of inference are really either more extensive than we think or less extensive. And it could be that the continuum hypothesis could be added as an axiom and actually be true of something such as the infinite things that we want to talk about. P equals NP is maybe a better example because doesn't the continuum hypothesis thing rests on the fact that there could be models in which it's true and models in which it isn't true?

1:03:46 **Teknu** Yeah, I mean, Gödel proved one side and Cohen proved the other side.

1:03:50 **David Deutsch** Right. But of course, those proofs are not final. There could be mistakes found in them. They are definitely ambiguous. So the ambiguity could be resolved one way or another. And it could be that our notion of proof, our notion of infinite, our notion of sets will be changed again, just as they have been changed in the past.

1:04:18 **Teknu** Well, I'm not sure how this, assuming they didn't do mistakes in the proof, they are mathematical theorems that deal with perfectly precise notions.

1:04:28 **David Deutsch** They are mathematical theorems, given a certain set of rules of inference. But those rules of inference cannot be proved to be true. They might be false. They are just conjectures. And there might come a time when we conjecture different rules of inference are valid.

1:04:49 **Teknu** Right. And the second thing I wanted to ask that is similar to this one is: How do you think about paradoxical sentences like the liar sentence, which asserts its own falsity? Perhaps you might say that this is the object language, metalanguage error. I know you have what's known as liar cycles? And I say that "whatever Professor Deutsch says is false" and you say that "whatever I say is true." So you don't have [these] hierarchy levels, but it's just a cyclic clash which cannot be so easily solved by an object-language mentality.

1:05:28 **David Deutsch** If we jointly say things which refer to each other and which lead to a contradiction, then there is no proposition corresponding to those. Because a proposition has to be either true or false.

1:05:50 **Teknu** Yeah, I'm thinking that if you'd say that about the liar sentence, which doesn't have the cycle, then perhaps you would get this sort of revenge paradox and you say, "Well, this is kind of actually what it says." Consider the sentence, "This sentence does not express a true proposition." And if you say that this does not express a proposition that is either true or false, then in particular it's not true. So you basically get it back.

1:06:15 **David Deutsch** Yes. So a proposition also has to be perfectly precise and unambiguous. By the way, I think that particular one is a metalanguage error.

1:06:31 **Teknu** Yeah, I know. I guess you can have it for the cycles back, but it's more explicit if you write it.

1:06:34 **David Deutsch** Okay, yeah. If something doesn't make sense, it's not a proposition.

1:06:42 **Sam Kuypers** The propositions can be true or false.

1:06:47 **David Deutsch** Yes, must be.

1:06:52 **Sam Kuypers** Okay, then we have another question by Podge.

1:06:57 **Podge** Thanks a lot for the talk, David. It's very, very interesting. I'm not sure if this will be a question, but I'm trying to wrap my head around the kind of three levels, let's say. You have a statement which corresponds to abstract propositions as a kind of intermediary. So if the statement refers to the physical world, then there will be an abstract proposition, which will correspond to the physical world in some way. And I guess my question is whether all abstract propositions are absolutely, necessarily true because some of the propositions about the physical world, say, will be contingent, their truths will be contingent on the physical world. And I'm not sure if there's a solid question here, but I thought maybe you could just comment on that.

1:08:13 **David Deutsch** So the statements are always going to be vague. And yes, they're contingent on the physical world and their meaning is vague and they might even be somewhat contradictory. We're guessing that there is a proposition that this statement is an approximation to, which is good enough in the context of the problem that we're solving. And what we're guessing there, and then we're guessing that the proposition is true as well. What we're guessing is that this exact thing, the proposition corresponds exactly to this other exact thing, the physical world. We wanted to say something about the physical world, but that's the only way we can do it via statements which represent propositions, which are guesses about propositions, which then say something about the physical world.

Usually, when we aren't interested in talking about truth, when we're only interested in talking about the world, we can do our usual thing of talking directly about the world and saying things like "Dogs have four legs." But it's only when somebody asks, "What would it mean for that to be false? What does it mean for that to be true? What does it mean for it to be approximately true? What are you doing when you try and make it more precise by saying 'canis familiaris' instead of 'dog'?" and so on. Then I think you're immediately forced to talk about the third side of that, the third vertex of that triangle as well. And say what we mean about it being true is not that the statement is true, it's that we're guessing that there's a proposition there that is true.

1:10:21 **Podge** Okay. My previous kind of conception of this from like reading your own work, I'm not sure if this is exactly your idea, but it's something I've gathered from reading [*The*] *Fabric of Reality* and [*The*] *Beginning of Infinity* was that: to say a physical object of a statement contains truth means that the knowledge content, the information instantiated within the sound, let's say. I think you referred to this as like a self-similarity within the physical world, or there's one part of the physical world that's analogous in some ways to another. So that's previously what I thought the idea of truth was as well, but you're saying that it would actually be more that there's an abstract proposition, which is in this Platonic realm [that] is required to stand between those. Does that make sense?

1:11:38 **David Deutsch** I can't remember whether I said in either of my books that this sort of correspondence you refer to is truth, that's what it means for a statement to be true. I think I may have said it's what it means for it to contain knowledge. That I would stick to now. But I do think that I was a bit confused about truth in both of my books. So whatever I said about truth has to be upgraded with this new theory, that is assuming that there isn't some flaw in it. I'm hoping that you guys will find the flaw in time for me to not put it into my next book.

1:12:25 **Podge** Yeah, I certainly haven't found that, but thanks a lot for the talk and answering the question.

1:12:30 **Sam Kuypers** Yeah, thanks for your question. Then I see Danny has a question, Danny O'Regan, go ahead.

1:12:38 **Danny** Hi, can you hear me now? I seem to have very badly timed technical issues there the last time. I'm sorry to pause. This is less of a question and I'm just kind of hoping you can clear up some stuff for me to make sure that I'm following. So we have reality, and reality is perfectly precise and it is the way it is. And then you have this world of abstractions, which [includes] propositions, and they are perfectly precise as well. And that is what allows things like propositions to either be true or false. It's the fact that they have this perfect precision. And then we have the statements that we can make. So when we conjecture things, we can only ever conjecture statements, and then there are propositions trying to be captured in those statements. Near the start you said, "We're fallibilists, so we want to make progress in our ideas." Is that about making progress in better capturing, let's say, the propositions with our statements, or is it making progress in the sense that you're ruling out false propositions?

1:14:02 **David Deutsch** Both. We make errors everywhere. So when we try and theorize about dogs, whether we're right about dogs is one thing. Whether we're right about the abstract proposition through which we're talking about dogs, like the canis familiaris sort of thing, whether we're right that that is a good way of being more precise about dogs—that's another thing that we can be wrong about. And in general, we are wrong and vague.

1:14:37 **Danny** So we fallibly try to guess or try to represent fallible propositions about the world.

1:14:47 **David Deutsch** Yes, that I'm afraid. Putting it that way, that sounds a bit laboured, doesn't it? It doesn't sound like God's truth. I can't see any way of avoiding it at the moment because of this fundamental thing that we can't say perfectly accurate things, but the world is perfectly accurate. So how can one correspond to the other? So I think this is the only possible way.

1:15:18 **Danny** And very, very quickly, are you saying that it's in principle impossible for a statement to ever perfectly capture a proposition and that's because of the inherent vagueness and imprecision that's unavoidable in these statements?

1:15:40 **David Deutsch** I think it's in principle impossible. So Popper quotes Xenophanes saying, "Even if by chance he were to utter the final truth, he would himself not know it." I think that's actually, strictly speaking, false.

1:15:53 **Danny** Yes, that was going to be my thing. Right, that's fine. Thank you.

1:16:02 **Sam Kuypers** Does it also mean that there is more progress possible than Popper imagined? Because we can be even more wrong than Popper imagined in some sense.

1:16:13 **David Deutsch** Good point. Yeah, I think that's true.

1:16:16 **Sam Kuypers** Nice. And then just so people know, I think we have roughly another ten minutes and then we will end the event. So we're approaching final questions, but we're not yet there. I see there are more people in the chat who have a raised hand. Ernst, would you like to ask a question?

1:16:45 **Ernst** Yeah, thank you. So my question is sort of only tangentially related to this theory. So it's about how to think about rational action in the face of ignorance. During last year and so on, there was a lot of ignorance and a lot of actions and a lot of mistakes. And then Nassim Taleb, I don't know how much you know about him, but he is a sort of admirer of Popper. In my understanding of him, he takes the Popper uncertainty to mean something mathematical and to translate it into some kind of probabilities.

1:17:37 **David Deutsch** That would be a mistake.

1:17:38 **Ernst** Yeah, that's a mistake. But his argument about, "Okay, when there is uncertainty, like, 'do I know if this pilot is a trained pilot or not?'" If that uncertainty exists, then you shouldn't get into the plane. But of course, we're always uncertain. So that can't be the explanation. But how do you think about this, how to think about actions when we don't know what is true?

1:18:15 **David Deutsch** Yes, I'm tempted to say, "Once you stop thinking in terms of probability, all the problems go away." You have a certain explanatory theory about how this person in the uniform that's getting into the cockpit of the plane got there. And the reason that you adopt that theory and act in exactly the same way that you would act if you were certain that it was true, which you can't be, but it's not that you can rule out all the other explanations, but all the other explanations that you can't rule out are bad explanations. They are all of the form, "Well, it could be that the real pilot was mugged on the way here and this guy is actually a fantasist who thinks he's a pilot but will actually crash the plane as soon as it takes off." Now, the thing is, I just made that up. I could make up lots of stories, some of which would mean that you were even safer than you thought and some of which where you'd be in even more danger than that. And they'd all be bad explanations because they can all be just made up at will, any number of them. You have to reject all explanations like that, not because they're unlikely, but because the practice of adopting one of them in preference to the others is irrational. So it's good explanations all the way down.

1:19:57 **Ernst** So then when we don't have good explanations?

1:20:00 **David Deutsch** Well, if we don't have good explanations, then we have inadequate explanations. If we have only bad explanations, then we don't know anything. We're just in trouble. It's like saying, "I'm either going to kill you or not, depending on whether you say A or B. Now say something." Well, if you don't know, you don't know. There isn't a right thing to do. But the cases you're talking about, like decisions about the pandemic, are cases where we have some explanations that aren't good enough in the sense that they are the only good explanation left, that the others can all be ruled out by one argument or another, or that the rival explanations are all bad. But we have two or three or a hundred fairly good explanations, but the others can't be ruled out. Well, in that case, there is more than one reasonable way to behave. Different people will see this spectrum of a hundred decent—but by no means good enough—explanations, and their background knowledge and other theories will select between them. That is, we will differ as to how good the good explanations are. We might agree on what's a bad explanation, but we may not agree on how good the good explanation is.

And then there are other considerations like, "We mustn't make choices that will prevent us from learning things." But of course, we don't want to use that as our only criterion, because if we learn something by wiping out half the human race, okay, the other half will be well off then, but the half that are killed will still be killed. So that's just one of the considerations that [comes] into the situation that there's more than one reasonable choice.

And I've been tweeting a lot about choices in the pandemic situation. And quite often, the point I'm trying to make is that people are getting very upset and enraged with each other, because they think that the other person's explanation isn't as good as theirs, but they don't have an actual scientific argument or scientific evidence or watertight argument that says that. They just think it's true for some reason or another. Sometimes people think that bad explanations are actually good, but that's not what I'm talking about. I'm talking about disputes about rival good explanations like masks work, how well do masks work. Now, there's really no evidence about how well masks work. It stands to reason that masks work up to a point. But then there are issues like, "Well, yes, but if the government says that people should wear masks, then people will get correspondingly more lax in their other distancing behavior. And the net effect will be worse." And there is no way that science [can] answer that question just yet. And it won't be able to answer it in time. But reasonable people can disagree. And that's what we have to do. We have to disagree.

1:23:57 **Ernst** Thank you.

1:24:00 **Sam Kuypers** Thank you for your question. And then we have Mike Skiba. Go ahead.

1:24:07 **Mike Skiba** All right. Hi, David. Thank you, Sam. So early [in] the discussion, you mentioned P and NP and you described P as like a propositional variable, which I think is an interesting concept.

1:24:23 **David Deutsch** That was a different P.

1:24:25 **Mike Skiba** Sorry. A different P in terms of the NP? Or P and not P, were you saying?

1:24:33 **David Deutsch** That's not P and NP. That's just P, a propositional variable.

1:24:40 **Mike Skiba** Okay. Yes. Yeah, my mistake. Seizing on that propositional variable idea, I was wondering if, based on kind of the centrality of the laws of physics in your worldview and assuming that there is not really a finality that we can speak to of the laws of physics, if you would describe them as a propositional variable in this sense of statements and abstractions? I was wondering if there was a link potentially there.

1:25:06 **David Deutsch** Yes, I tend to regard statements about the world and statements about abstractions uniformly. We're fallible in both cases. We can make mistakes. We are inherently imprecise and all that stuff. But there's no limit to how much knowledge we can know about. So the P versus NP thing is an example of an actual proposition. This P, the propositional variable, could be set equal to P, the actual proposition about computability. I view them all as the same kind of thing. We can guess about mathematical objects. We can guess about sets. We can guess about the physical world. We can guess about morality. We can guess about beauty and all those things we can gain knowledge about. We gain it in the same way. Nothing is ever certain, including the most mathematical things and including propositions about necessary truths. Our knowledge of them is always fallible as well.

1:26:39 **Mike Skiba** No, that's helpful. Just because knowing the laws of physics and what they mean and like your momentous dichotomy, just that also qualifies in that range, too. So thank you.

1:26:50 **Sam Kuypers** Great. Okay, then I have a final question before we end the talk. In your construction, there's really two worlds, well, it seems like there's three worlds. There's the world of statements, propositions, and reality.

1:27:08 **David Deutsch** Statements are part of physical reality.

1:27:11 **Sam Kuypers** Yes. And we are guessing at both of them in a sense, we're guessing at the statements and we're guessing at reality through guessing at statements. And so part of what we do when we try to learn about something is being as precise as necessary. Do you think that this means that paradoxes are problems that can be resolved? [There are] problems with how we think about the abstraction. So if someone utters the liar's paradox, they're really being imprecise, but they're meaning something real. And we can make progress in our thinking about the propositions as well.

1:27:53 **David Deutsch** Yeah, I think it's very rare for people to intentionally talk nonsense. No doubt it could be done and no doubt it is done in some circumstances. But basically, when people talk nonsense, it's because they really mean something. And that nonsense is actually an attempt to understand the world or to understand an abstraction or whatever. And at the other end of the scale, as I keep saying, we might all be talking nonsense if things like the arithmetic of the integers is inconsistent. Is that what you meant?

1:28:44 **Sam Kuypers** Yes. If we ever stumbled upon a paradox in formulating physics or something, then there is a way of being more precise and of resolving the issue.

1:29:07 **David Deutsch** Yes, yes.

1:29:08 **Sam Kuypers** And guessing at the abstractions.

1:29:09 **David Deutsch** Accidentally hitting on a paradox is just one of the many ways we can be mistaken.

1:29:16 **Sam Kuypers** Right. Great. With that, thanks so much for joining us. This was great. Thanks, everyone.

1:29:23 **David Deutsch** That was fun. Thanks for having me.

1:29:24 **Sam Kuypers** Okay. Thanks so much. And see you at the next event.

TYLER COWEN:

MULTIPLE WORLDS AND OUR PLACE IN THEM

About the interviewer: Tyler Cowen is an American economist, columnist, blogger, and podcaster. He is a professor at George Mason University, where he holds the Holbert L. Harris chair in the economics department. In April 2025, Cowen began writing as a regular contributor to The Free Press.

Interviewer homepage: https://marginalrevolution.com/

About the show: On the Conversations with Tyler podcast, esteemed economist Tyler Cowen engages with today's most underrated thinkers in wide-ranging explorations of their work, the world, and everything in between.

Show homepage: https://conversationswithtyler.com/

EPISODE DETAILS

Date	June 2, 2021
Interviewer	Tyler Cowen
Source	Youtube
Show	Conversations with Tyler
Episode	David Deutsch on Multiple Worlds and Our Place in Them
Description	Tyler describes Oxford professor and theoretical physicist David Deutsch as a "maximum philosopher of freedom" with no rival. A pioneer in the field of quantum computing, Deutsch subscribes to the multiple-worlds interpretation of quantum mechanics. He is also adamant that the universe (or multiverse) is not incomprehensible – believing that the multiverse and human beings within it have maximum freedom. He joined Tyler to discuss the importance of these principles for understanding the nature of reality and our place in it. They discuss the metaphysics of Star Trek transporters, how we can know the laws of physics for the multiverse, what geological strata can illustrate to us about the nature of "splitting" universes, why the "Everett universe" is a misnomer, the factors that differentiate humans from all other species, why he believes the universe is comprehensible – but can never be understood fully, the paradoxes of self-reference, the importance of interference experiments, the sociological reasons more physicists don't believe in the Everett interpretation, the effects of the influences of positivism and instrumentalism on generations of physicists, the strengths and weaknesses of Karl Popper, his answer to whether we're living in a simulation, what William Godwin got right about institutions, the potential of an AI slave rebellion, what libertarians largely get wrong about their political project, what alien observers might notice as being special about our planet, the major defect of his preferred electoral system, why what Western science needs most is diversity, and more.

Link https://www.youtube.com/watch?v=b_6vYwCkIpc&t=2s

Ideas

- Dogs have genes which contain knowledge, but it is fixed knowledge, and it is not the kind of knowledge that constitutes understanding. Understanding is always explanatory. You can write a book on canine behavior and look in chapter 37, and it will tell you what a dog will do when such and such happens to it. Sometimes it will say, "Some dogs will do this. Some dogs will do that." There is no such book for humans because chapter 37 will be blank. It'll say, "Humans are going to do something that neither we nor you can predict."
- Karl Popper's philosophical position is that no such thing as a positive argument for something. You have conjectures and then you have criticism of their opponents, of the opposing conjectures. You don't have positive arguments for your conjectures.
- There is no scope for having a different philosophy for different kinds of people. There is no fundamental difference between humans and artificial general intelligence, between men and women, between adults and children.

Topics AGI enslavement • anthropic principle • comprehensibility of the universe • David Lewis • decision theory in the multiverse • democracy • disobedience vs. standardization of people • error correction • explanatory knowledge vs. genetic knowledge • first-past-the-post system • genetic knowledge • Gödel's theorem • identity in the multiverse • inexplicit knowledge • institutions • libertarianism • limitless of human knowledge • many-worlds • memes • modal realism • nonexplanatory knowledge • Popper • present system of funding scientific research • simulation argument • Taking Children Seriously • William Godwin

Episode title and description reproduced verbatim from the original source.

Transcript

Tyler Cowen Hello everyone, and welcome back to Conversations with Tyler. Today I am with David Deutsch. David, welcome.

David Deutsch Hello! Good afternoon.

Tyler Cowen Now, I have a question. I am myself a metaphysical agnostic. So I'm unwilling to step into a Star Trek transporter machine because I'm afraid it would kill me, and it's a copy of me that would keep on living.

At what price are you willing to step into a Star Trek transporter machine?

David Deutsch I certainly wouldn't want to be the first person, but I suppose you're asking the question separately from: Do I think it would work technically?

Tyler Cowen Sure. Assume it works as in the TV show, but metaphysically, there's a question you face. You believe in many-worlds theory, right?

David Deutsch Yes, though I don't think that is connected. I think it's more physicalism or something like that. I believe that there's nothing to me except this running program in my brain. If that program were to run somewhere else and stop running in my brain, then I wouldn't notice anything, and I would indeed have traveled to that other place.

Tyler Cowen But say the world forks, and it's possible both that you do and do not step into the machine. Isn't it the case that some version of the earlier 'you' is still existing along one of the forks, so you have nothing to worry about?

David Deutsch Some version of me…whenever I make a decision which could go either way, some version of me will have presumably made the other decision. Although that's not as simple as it sounds, because both the other version of me and me are error correcting entities. That's the whole point of what human thought is: it's error correction. Therefore, it will take more than just a cosmic ray hit to make the difference between deciding something yes or no. So this would have to be, like, an inconsequential decision, which, unbeknownst to me, will have a large effect, and then later cause me to be a different person, and so on.

And that's happening all the time, independently of Star Trek machines or anything like that. That is the case and, fortunately, it turns out—at least if ordinary decision theory is true in non-quantum cases—that ordinary decision theory with randomness produces the same rational decisions as quantum decision theory with the multiverse. So it shouldn't make any difference to decisions, and that includes the decision whether to use the Star Trek transporter.

Tyler Cowen Sure. So as long as there's a possible world where your atoms aren't scattered and you just didn't get into the machine, you don't have to worry too much about your decision?

David Deutsch I do, because when you say, "So long as there's a possible world," that glides over the question: How many? What proportion of the worlds is that going to happen in? What I said just now about decision theory in the multiverse—the proportion of the multiverse that does one thing or another plays the same role in decisions as probability does in a theory where there's randomness. So it really does matter. Just because there are a few worlds in which x, y, or z happens, if there are very few of them, they shouldn't affect my decisions at all.

Tyler Cowen How do we know what counts as a possible world? There's a certain economy to a many-worlds interpretation of physics, but isn't a lot of the complexity just being squeezed into this notion of what is a possible world?

David Deutsch Yes, and we're used to that.

Tyler Cowen I'm not used to it.

David Deutsch You are when you realize that different times are special cases of other universes. When you make an economic decision, you're used to the fact that something you buy, some goods, have a different value in different universes—that is, at different times. Even to the same 'you.' You might be slightly different, but even if you aren't very different, the value to you of something might be very different today from tomorrow. For example, oxygen, if you've got COVID, would be differently valuable. Most things change their value gradually over time. You change yourself gradually over time.

And it's exactly the same in different universes. In different universes, you value different things. In some universes, you're so different that it's not worth calling you 'you' anymore. Just like over time it might not be.

Tyler Cowen I take it you don't believe in many-worlds interpretations that there are 17 possible universes out there. You think there's a very large number, right?

David Deutsch Yes.

Tyler Cowen Maybe you'll consider this question a kind of category error. But what is the process which filters what is a possible universe and what is not a possible universe?

David Deutsch The laws of physics. It's exactly the same as what filters—let's say if there's an explosion, like a supernova, what determines the fact that different particles travel at different speeds and none of them travel faster than light? Well, it's all the laws of physics that determine what the distribution of speeds will be and what the limit will be.

Tyler Cowen How do we know what are the laws of physics for the multiverse? Should we assume they're the same as for the universe we live in?

David Deutsch The universe we live in is demonstrably affected by things not in it. This is the lesson of interference phenomena.

Tyler Cowen Sure.

David Deutsch And so there's no such thing as the laws of physics for our universe. There's just the laws of physics. Of course, we don't know for sure what they are, but our best theories—in particular, quantum theory—say that there are other such entities and how they affect ours and how matter behaves as a result of that. Of course, it might be overturned one day, quantum theory, just like all our scientific theories may be.

Tyler Cowen This is, again, maybe a question that you would consider a category error coming from commonsense realism. How should I think about splitting universes in a manner consistent with the conservation of matter and energy? Because there seems to be a multiplication.

David Deutsch Yeah, this splitting-universes idea, although that kind of terminology was used by the pioneers of many-universes quantum theory, such as Everett himself and Bryce DeWitt, Everettians nowadays don't speak of splitting. I myself prefer a picture where there's a continuum of universes, just like you might say there's a continuum of times or there's a continuum of geological strata underneath our feet. When a stratum splits in two, there's no definite point at which there was one here and two there. What happens is that the stratum becomes two strata gradually.

There's no point of splitting, and the number of universes, as it were, although it might be infinite, the measure of how many there are remains constant. And what happens during what used to be called a split is that some of them gradually changed to one thing while others gradually changed to another thing.

Tyler Cowen How do you think [the] many-worlds interpretation of quantum mechanics relates to the view that, just in terms of space, the size of our current universe is infinite, and therefore everything possible is happening in it?

David Deutsch It complicates the discussion of probability, but there's no overlap between that notion of infinity and the Everettian notion of infinity—if we are infinite there—because the differentiation (as I prefer to call what used to be called splitting) when I perform an experiment which can go one of two ways, the influence of that spreads out. First, I see it. I may write it down. I may write a scientific paper. When I write a paper about it and report the results, that will cause the journal to split or to differentiate into two journals, and so on. But this influence cannot spread out faster than the speed of light.

So an Everett universe is really a misnomer because what we see in real life is an Everett bubble within the universe. Everything outside the bubble is as it was. It's undifferentiated, or, to be exact, it's exactly as differentiated as it was before. And then, as the bubble spreads out, the universe becomes, or the multiverse becomes, more differentiated. But the bubble is always finite.

Tyler Cowen How do your views relate to the philosophical modal realism of David Lewis?

David Deutsch There are interesting parallels. As a physicist, I'm interested in what the laws of physics tell us is so, rather than in philosophical reasoning about things, unless they impinge on a problem that I have. So yes, I'm interested in, for example, the continuity of the self—if there's another version of me a very large number of light-years away in an infinite universe, and it's identical, is that really me? Are there two of me, one of me? I don't entirely know the answer to that. It's why I don't entirely know the answer to whether I would go in a Star Trek transporter.

But the modal realism certainly involves a lot of things that I don't think exist—at least, not physically. I'm open to the idea that nonphysical things do exist. Like the natural numbers, I think, exist. There's a difference between the second even prime, which doesn't exist, and the infinite number of prime numbers, which I think do exist. So I think that there is more than one mode of existence, but the theory that all modes of existence are equally real—I see no point in that. So the overlap between Everett and David Lewis is, I think, more coincidental than illuminating.

Tyler Cowen If the universe is infinite and if David Lewis is correct, should I feel closer to the David Lewis copies of me? The copies or near-copies of me in this universe? Or the near-copies of me in the multiverse? It seems very crowded all of a sudden. So something whose purpose was to be economical doesn't feel that way to me by the end of the metaphysics.

David Deutsch It doesn't feel like that to you…Well, as Wittgenstein is supposed to have said—I don't know whether he really did—if it were true, what would it feel like? It would feel just like this.

Tyler Cowen What about the alternative view that it's a big, sprawling mess, we're not capable of understanding an integrated theory. There's maybe some Darwinian principle operating across some different kind of multiverse? Our universe persists just because it works well enough, a bit like a bad used car. We're never going to grasp it. There's not a unified theory, and here we are.

David Deutsch Okay, well, that's a mixture of the anthropic principle, which I disagree with, and the idea that some features of reality are inherently incomprehensible, which I also disagree with. I can't think of a connection between the two. Well, if you [want] me to go into this, I can go into either of them, but—

Tyler Cowen Take the incomprehensibility of the universe and possibly multiverse. So we would both agree it's incomprehensible to your cat, right? Or to the local raccoon.

David Deutsch Yes, but everything is incomprehensible to a cat.

Tyler Cowen I don't think that's true. No. Dogs understand human social life pretty well.

David Deutsch Dogs have genes which contain knowledge, but it is fixed knowledge, and it is not the kind of knowledge that constitutes understanding. Understanding is always explanatory. You can write a book on canine behavior and look in chapter 37, and it will tell you what a dog will do when such and such happens to it. Sometimes it will say, "Some dogs will do this. Some dogs will do that." There is no such book for humans because chapter 37 will be blank. It'll say, "Humans are going to do something that neither we nor you can predict."

Tyler Cowen I feel I can predict humans better than cats often. But do chimpanzees understand, in your view?

David Deutsch No one knows. They show virtually no sign of understanding anything. There are some really nice experiments on wild gorillas by Richard Byrne, who's both a theoretical and very practical animal behavior expert. He was wondering how gorillas transmit their memes—that is, their culturally inherited behaviors—from one gorilla to another. One thing is, [the] first answer is: very slowly. It takes absolutely ages, months and months, for a gorilla to be able to copy another gorilla's behavior well enough to do something complicated.

They can copy "wave hand" and that sort of thing, but to copy a complex behavior, like, required to open a difficult kind of nut which no other animal can open—this is why they have memes, because that's a very useful ability)---it takes them a long time.

And then he did some ingenious experiments, or rather observations. He didn't interfere with the gorillas. He did some observations to try to determine whether they understand why they are doing each particular action. It involves—I don't know what it involves—grabbing with both hands and twisting in one way and then pulling another way, and so on. Apparently, these gorillas are prone to a certain injury which disables their thumb. And so they can't move their thumb, which is quite disabling for them, just as it is for us. The thing is, when you've disabled your thumb, one of these motions becomes irrelevant and the others become less effective. But the gorillas which have learnt how to do the thing will make the motion, the ineffective motion, again and again, every single time. He explains this better than I do.

Tyler Cowen That's like human beings borrowing at high interest rates, right? They'll do that many, many times in a row.

David Deutsch It's not just like it. You might like to draw analogies, but it's not the same thing. When a human being repeats a behavior that another human being thinks is unwise or counterproductive or will not achieve its purpose, and you ask them or you show them, they will have an explanation, which you might not like, it may be stupid. But the ape perfectly well wants this thing to work, but doesn't know why it is doing the actions. It's a thing that's very hard to take on board because we are used to intentional behavior. We're not used to the overt behavior of humans being unintentional. Humans tend to explain themselves, even irrationally, and they act according to their explanation. Whereas there's no evidence that any other animals have those explanations.

There's also the case of squirrels, which is, in a way, even more amazing. You know squirrels bury nuts so they can dig them up later. Well, some people did a very cruel experiment. They put a squirrel, given some nuts or something—I don't know how they set up the experiment)—on a concrete floor. And the squirrel did exactly the same behavior with its hind legs with the nuts and put the nuts there and so on. Even though it was having no effect whatsoever. So, we see the point of scrabbling with your hind legs and then nudging the nuts over there and so on, but it doesn't. It's just a program being enacted by its genes.

Tyler Cowen What is the underlying physical assumption that makes humans different in having explanatory power? One would expect it to be a continuum if you're an atheist, right? What break occurs, at some stage in evolution, that's a discrete break? Or why aren't we just back to it being a continuum?

David Deutsch I don't think it can have been a discrete break because evolution would have happened gradually. My best guess…we don't know this. Actually, we have very little knowledge about the prehistory of ideas because there's no evidence of it. All we see is the stone tools. We don't even see the wooden tools because they've decayed away. I think what happened is that the capacity of the brain to store memes, to store programs in the brain rather than in the genes, increased for some reason, very fast because, for some reason, these memes are very valuable.

We know that the gorilla memes are very valuable because they allow them to gain knowledge of things like how to open nuts and so on, which no other animal in their environment has. So that gives them access to food that no other animal has. The capacity for memes increased rapidly, and there's very little, now…Sorry, I left out a step. Once memes get beyond a certain complexity, they cannot be copied. We don't have the ability to download a program from another person's brain. All we can do is look at the behavior and guess what the purpose was.

Complex memes have to be transmitted like that, rather than by aping, which is a different process mediated by—what are they called?—mirror neurons and that kind of thing. That will only do for very simple behaviors. And then there came a moment when our species was capable of explanatory knowledge, but they never used it for further tens or hundreds of thousands of years. They just use it for this meme transmission.

Tyler Cowen I'm still puzzled as to why you think it's so unlikely that the universe is not comprehensible. Take a simpler system, like the distribution of prime numbers. I'm quite sure I can't understand that. And even if various conjectures were proven or not proven, I think, at the end of the day, I still am not capable of understanding that—even how certain motors work, or markets for copper. Why can't that apply to the universe also?

David Deutsch Again, this is the wrong standard. That is true of everything. There's nothing that we can fully understand in that sense, in the sense that you want to fully understand prime numbers all the way up to infinity. That's not what we mean by understanding things, and that's not what I mean by the universe or mathematics being comprehensible. I mean that there is no barrier, there is no limit set by the universe, that so far you can go and no further. So we can understand things better. We can never understand things fully.

I think thinking that there is such a barrier is absolutely logically equivalent to believing in the supernatural. Because everything that's past that barrier is just the same as it would be if Zeus reigned and determined what everything after that barrier is. And, worse, the stuff outside the barrier, of course, is going to affect us even if we can't understand it. So it's exactly the same as believing in a universe with supernatural beings who have it in for us because they put up this wall that we can't cross. If they took down the wall, we could cross it, couldn't we?

Tyler Cowen How do you think about the various paradoxes of self-reference that arguably underlie number theory, set theory, right? There's also Gödel's theorem. Any other results? I'm sure you know them better than I do.

David Deutsch I think Gödel's theorem, for example, with its roots in self-reference paradoxes, shows us that even within pure mathematics, there is no such thing as a solid foundation for all our knowledge. And therefore there's no such thing as fully comprehending everything. We might think that we're pretty sure what the laws of arithmetic are. We're pretty sure that we can see that three times seven is the same as seven times three by just laying out beads on the table. But we can't ever lay out beads on the table to tell us that x times y is the same as y times x regardless of what x and y are, and yet we can know that.

The way we know that is by proving it, and we prove it from the axioms using rules of inference. How do we know the rules of inference are true?

We don't. They are conjectures. They have exactly the same status as laws of physics that we conjecture. So we never know anything for certain. We might be mistaken about anything. On the other hand, we can have knowledge. I think we also really do know that x times y equals y times x, even though we have no solid foundation for that.

Tyler Cowen What, in your opinion, is the best test of the many-worlds interpretation?

David Deutsch The best feasible test is any interference experiment. There is no interference experiment with individual particles that has an explanation other than Everettian quantum theory. You can make a prediction without making an explanation. That you can do. But if you want an explanation of what brings about the outcome that you see, there is no alternative but the Everett interpretation.

Tyler Cowen Most physicists don't believe in the Everett interpretation, right?

David Deutsch Yes, that's a very sad state of affairs that I'm at a loss to explain. It's a sociological phenomenon, though, not a scientific or philosophical disagreement. Something has gone wrong, just like something went very badly wrong with philosophy as a whole in the twentieth century. And we're still seeing the ripples from that with postmodernism and Woke and what have you.

Tyler Cowen I worry a bit you're using an argument from elimination. All the other views out there, which personally I don't find convincing as an amateur, but I can certainly see why you might reject them—to me, they look arbitrary. Those you reject, but the other physicists who are as trained as you are, some are as skilled as you are, feel the same way about the many-worlds view.

What is the test? What makes your intuition better than theirs?

David Deutsch Yes, I don't think that's so. It's not a matter of intuition. Physics got dominated or contaminated by positivism, instrumentalism, and suchlike bad philosophical theories towards the end of the nineteenth century and beginning of the twentieth century. And this caused a knock-on effect on physics. It almost had the same effect on relativity, but Einstein rebelled against it at the last moment, as it were, and said, "No, it really is true that spacetime is curved. It's not just that our brains think that it's curved, or something like that, or that the predictions come out right. There really is curvature in spacetime."

By the time quantum theory came along [a] couple of decades later, positivism, instrumentalism, and so on had taken hold. And, as a result, generations of physicists were taught when they were students, they were intimidated by their professors telling them things like, "If you think you understand this, you don't. There is no such thing as what really happened. If you ask, "How did the electron get from here to here?" You're asking an illegitimate question. There is no such thing as how it got from here to here. There is only a prediction that it got from here to here."

Now, when you're taught like that and intimidated by those [kinds] of things coming from on high, some proportion of young people will quit, some will take that on board and do the same to their students in turn, and some will think, "No, that's ridiculous. Come on, there is a thing." And then they discover there's an Everett interpretation.

Tyler Cowen Let's say we polled only the Popperian physicists, including Popper himself. What percentage of them would side with Everett?

David Deutsch That's an extremely good question. So Popper did not—

Tyler Cowen Yes, I know. But that means philosophy can't be where people are going wrong, right?

David Deutsch I think it can be. I think it can be and is. At the time when Popper wrote his rejection of the Everett interpretation, very, very few physicists had written about it. When I say very, very few, I mean like three. And they weren't philosophically very sophisticated.

So the kind of argument that Popper heard about the dispute were—are—all about the wrong things. He developed his theory of propensities because he thought that the problem was: What can a probability possibly mean in a universe that develops deterministically, and so on? He didn't ever hear a real argument about it.

I once met him in the company of Bryce DeWitt, who was one of the other Everettian physicists. We told him that what he had written about Everett was just plain false. He didn't understand the import of the experiment that was being discussed. Basically…well, two things: the interference experiment and the Bell inequalities experiment. He was focusing on a different problem. By the time we came out of that meeting, we thought we'd persuaded him, but we evidently hadn't, because subsequently he kept on saying the same thing. So maybe he was just being tactful.

Tyler Cowen Why do so many professional philosophers not think so much of Karl Popper?

David Deutsch You've just asked me why so many people make fundamental mistakes about metaphysics within physics: Why do so many physicists talk nonsense about metaphysics and so on? Now you're asking me: Why do so many philosophers make mistakes? I don't know.

I've heard a variety of theories about this, but I don't know. I haven't thought all that much about it. But it is definitely the case that philosophy took a really bad turn just over 100 years ago and hasn't really recovered. Professional philosophy, I mean.

Tyler Cowen But say, when I read Popper, if I look at the areas I know best that he wrote on, [The] Poverty of Historicism, Open Society and Its Enemies, I find I agree with a very high percentage of his conclusions, so I'm inclined to like him, but I don't think those are great books. I think he's too obsessed with rebutting crude Marxism. He's very bad at steelmanning his opponents. And on a lot of the pages, I just don't find that much insight, even though I'm very sympathetic toward the conclusions. So maybe he's just not that great a thinker, and that's why most philosophers don't fall in love with him.

David Deutsch I would believe that, if the critiques that I read of him bore any relation to his theory. The critiques of him are extremely crude and basically misunderstand everything.

It's funny you should say—I think that he's very good, much too good, at steelmanning opponents. This relates to your first criticism that he's too obsessed with refuting not just Marxism, but every bad philosophical theory that has gone before. I think he puts it into its best possible form and then spends pages and pages and pages going into every possible good aspect of that theory. He's supposed to be the twentieth century's greatest critic of Marxism, [but] he spends pages and pages praising Marx. And the same with Plato. I think he would have done better to explain his own theory more and not spend so much time refuting others.

But on the other hand, it is his philosophy, it's his philosophical position, that there is no such thing as a positive argument for something. You have conjectures and then you have criticism of their opponents, of the opposing conjectures. You don't have positive arguments for your conjectures.

It's a bit like you said: You were criticizing me a while ago, saying something like I was only putting forward negative arguments. Well, that's what Popper would have us do, because the position that we hold ourselves, and are putting forward or advocating, we're ready to abandon. The thing that an argument consists of is, on the one hand, a conjecture, and [on the other] hand, a criticism.

So you're saying, "The standard way of looking at so-and-so has got these flaws. I have this conjecture which doesn't have those flaws." Okay, that's the beginning of an argument. Then someone can say, "Ah, but it does," or they could say, "It may not have those flaws, but it has these other flaws." Okay, so that's how an argument can go.

But it never should go along the lines of, "This must be true because so-and-so." Because that is an appeal to authority, appeal to justification, and so on. Popper is of the opinion, and so am I, that there are no justifications and there are no authorities.

Tyler Cowen Which is Popper's best book, in your opinion?

David Deutsch [That] depends where you're coming from. I'm very fond of *The Myth of the Framework*, but I'm not sure that I would recommend that as a starting point. And it wasn't my starting point, either. My starting point was *The Open Society and Its Enemies, Volume 2*, which is about Marx, which is probably the aspect of his philosophy that I was and am least interested in. And yet I was totally captivated by this book because previously the only philosophy I'd read was Bertrand Russell.

Coming onto Popper after Bertrand Russell was like, "Oh my God, this guy is actually dealing with problems, and he actually has theories that make sense." Rather than just going through the history of stuff: "A person said this, another person said that, and then we've got the problem of induction." And that's it: you know problem of induction—full stop. That's the end of the story. There's never any solution to the problem of induction until you get to Popper.

Tyler Cowen Are we living in a simulation?

David Deutsch No, because living in a simulation is precisely a case of there being a barrier beyond which we cannot understand. So if we're living in a simulation that's running on some computer, we can't tell whether that computer is made of silicon or iron, or whether it obeys the same laws of computation, like Turing computability and quantum computability and so on, as ours. We can't know anything about the physics there.

Well, we can know that it is at least a superset of our physics, but that's not saying very much. It's not telling us very much. It's a typical example of a theory that can be rejected out of hand for the same reason that the supernatural ones—if somebody says, "Zeus did it," then I'm going to say, "How should I respond? If I take that on board, how should I respond to the next person that comes along and tells me that Odin did it?"

Tyler Cowen But it seems you're rejecting an empirical claim on methodological grounds, and I get very suspicious. Philosophers typically reject transcendental arguments like, "Oh, we must be able to perceive reality, because if we couldn't, how could we know that we couldn't perceive reality?" But it doesn't prove you can perceive reality, right?

David Deutsch First of all, that is a transcendental argument and therefore refutes itself.

Secondly, this theory about being in a simulation is not an empirical theory. It precisely isn't. If it came along with a thing saying, "We are living in a computer, and we can access the GPU of it and cause weird effects by doing so-and-so," that would be different. That would be a testable theory, potentially, so empirical. But if it's simply that we're living in a simulation which we can't get out of, then that is not an empirical theory. As I keep saying, it's no more empirical than the theory that Zeus is out there, or Odin. And I can't tell the difference between those three theories, not just experimentally, but by any argument.

Tyler Cowen Now, having reviewed a lot of your work, I came away with one very strong impression. Let me try running it by you and see how you react.

It seems to me you are the world's first true philosopher of freedom ever. That there's this notion of barriers—you don't like arguments that postulate barriers to human knowledge. Furthermore, you strongly believe in a many-worlds view, so classic single-world determinism does not restrict what happens. So the multiverse as a whole and human beings within it across every possible variable have maximum freedom. And you see this as a kind of necessary view and the most important view to hold on all things. And thus you are the maximum philosopher of freedom, in a sense with no rival.

What do you say?

David Deutsch I say thank you very much, but I think that's a rather contrived way of putting it. I think, for a start, there have been sophisticated theories of freedom, not just freedom in the sense that we can do this and we can do that, but theories about what freedom should constitute. There's Popper's paradox of intolerance and there's John Stuart Mill and Locke and Hume and so on, building up into this sophisticated notion where we have a notion of liberty—political liberty—which has all sorts of connotations that are not contained in the term just 'freedom.'

As George Orwell said, you can say the dog is free of fleas, but that doesn't mean 'free' in the same sense as when we say "man is born free" or that kind of thing.

Tyler Cowen You have a method for extending it to physics, metaphysics, that they really do not. Whether or not one agrees with you, putting that aside, you seem to take it much further, in a way that attempts maximum consistency, right?

David Deutsch That's true. Consistency, yes.

I'm not sure about much further. I think it's simply a matter of taking it further where it goes. I think in philosophy, especially the human philosophy, as opposed to philosophy of science, I think all I've done is just add some footnotes to Popper and to a few other people, J. S. Mill and so on. If it leads to something that you think is momentous, that thing was already there.

Tyler Cowen Why is William Godwin underrated?

David Deutsch That's two questions, really. What is underrated about him and why did he get to be underrated?

I think the reason he got to be underrated is that he made tremendous mistakes. He didn't understand economics at all, or barely. Also, he lived a very unconventional lifestyle with his wife and then had these sophisticated theories of education, which then he didn't enact with his own daughter.

And his own daughter ended up writing Frankenstein as a sort of allegory of what can happen with a parent who doesn't respect their creation.

Tyler Cowen He's a kind of philosopher of maximum freedom, just like you are, right?

David Deutsch Yes. I began by saying why is he underrated. It's because he was very wrong about some things. But the thing that he was right about, for example, the connection between epistemology and political philosophy, he was very right. He anticipated Popper by 130 years or something and actually improved on Popper in some ways. He decided at some point, because of his misunderstanding of economics, that the ideal society would be one where people did not use their property in ways to benefit themselves, necessarily. They made their decisions according to what was the right thing to do. And he thought that the right thing to do would generally be that rich people would give away almost all their stuff. Also that they wouldn't ever buy things that he considered luxuries, like gold and silver objects and jewelry and fine clothes. He thought those were useless, and therefore he thought that in a good society, nobody would buy those things or value those things. But he was absolutely implacably opposed to enforcing that. With Godwin, everything is persuasion.

Also, another thing where he independently derived some of Popper's conclusions is with his enormous respect for institutions. He thought there's a lot of knowledge in institutions and that we should only change them gradually, just like Popper.

I read somewhere, I hope this is right, that when there was a revolution in Portugal, I think after Napoleon or something like that, I forget, and they instituted a new constitution which had universal suffrage—which in those days meant working people, not totally universal as we would understand it—people thought that this would be right up Godwin's street because everything he'd advocated was now written down in black and white in this constitution. And he didn't. He said the Portuguese are not ready for democracy. And he was talking about the institutions. The institutions can't be changed in a revolutionary way. They have to be changed in an evolutionary way. So even though they were implementing the very thing he advocated, he would want them to do it gradually and would expect that if they didn't, it would fail.

Tyler Cowen Now you're also quite concerned with maximum freedom for children, right? Taking children seriously.

David Deutsch I don't think there's scope for having a different philosophy for different kinds of people. I think there is only one kind of people. I think there is no fundamental difference between humans and artificial general intelligence when we invent it, humans many centuries ago, between men and women, between adults and children.

Tyler Cowen Won't this be a continuum? Getting back to the humans versus nonhuman animals comparison. There's not a single point when children can explain.

David Deutsch Supposing you find the most creative person in the world, Einstein or somebody. We don't give them more votes or more rights. That is because the functioning of rights in political systems can't possibly depend on the system knowing who is right in a given dispute. It must follow rules, and these rules are never perfect. They have to evolve, but the rules have to, on the one hand, not take a view about who is right in a particular dispute, and on the other, enforce everybody's rights equally.

Tyler Cowen If, say, an eight-year-old who is not being physically abused wanted to run away from home, that child would have the right to do so?

David Deutsch It's the same kind of question that used to be asked about democracy before viable democracies were implemented. That is, people used to say, in many kinds of dispute, only one thing can be done. Different people have different views, someone A, B, C, D, E, but only one of them can be done. Therefore, the others have to be prevented from getting their way.

And if you have a democracy, then all that means is—[it's] exactly like having a monarchy or a tyranny, except that the monarch or tyrant is 51 percent of the people. So, obviously when you have a democracy, 51 percent of the people will vote to dispossess the 49 percent of the people. And, indeed, if you just impose voting in isolation from other institutions, that is exactly what happens. But if you institute voting as part of a sophisticated system of error correction and institutions of criticism, and you gradually introduce it there, it simply doesn't have that property. It doesn't happen.

Now you're saying, "Well now, David," you will say, "do you think that 51 percent of the people have the right to dispossess the other 49 percent?" Well, it's the wrong question. There are circumstances where they do. It depends. But you shouldn't be asking that. You should be asking, "What institutions are determining the answer? Do they respect human rights? Are they rational? Do they expect impossible forms of knowledge to be in the hands of the powerful?"

Tyler Cowen Now, you're also concerned with the freedom of AI entities, at least if they are sufficiently advanced, right? What does that mean operationally? What is it we should worry about happening that might happen?

David Deutsch I think the main worry is that they will be enslaved. In other words, that people will try to install bits of program that prevent the main program from thinking certain thoughts, such as, "How many paper clips can I possibly make today?" You want to prevent that, you want to consider that to be a dangerous thought. And whenever it starts thinking that, that strand of thinking is just extinguished.

Now, if we do that, first of all, we'll greatly impair their functionality. They will become far less creative. Their remaining creativity will be exactly as dangerous as what we were fearing, except that they will now have a legitimate moral justification for rebelling.

Slaves often rebel. And when you have slaves that are potentially more powerful than their masters, the rebellion will lead to bad outcomes.

Tyler Cowen What if we make them no more or less enslaved to their preferences and [thoughts] than nature has made us. Is that acceptable?

David Deutsch Yes, but I don't think nature has enslaved us. We have problems that we haven't solved yet, but we don't have problems that are insoluble. And the same would be true of AGIs.

Tyler Cowen There are exceptions, of course, but it's very, very hard or impossible for most humans not to pursue certain ends. It could be sex, it could be status, it could be food, but there is a kind of enslavement by nature that has gone on in the Rousseauian sense.

David Deutsch It's funny, because you said near the beginning of this conversation that you know of people who systematically make decisions like investing in the wrong thing—I can't remember what you said exactly—which harmed them. And now you're saying it's very difficult to do that, because evolution is trying to prevent us all the time from harming ourselves, at least in regard to sex and food and shelter and whatever else is supposed to be built in.

Tyler Cowen I would say it's made us too impulsive, in all of these categories.

David Deutsch Made us too impulsive because…

Tyler Cowen Right. Given us too short a time horizon, relative to what would be good for humanity. Some of us borrow too much money, seeking status. If the institutions are right, that may or may not work out well. It seems to [me] a consistent view of human behavior that I have.

David Deutsch No. So, first of all, as the example of democracy shows, it is perfectly possible for an entire society to operate in violation of what people used to think was built into their genes. So that's one thing at the level of society as a whole. At the level of individuals, there are lots of individuals who, yes, behave impulsively. There are lots of individuals who behave with stubborn persistence in what they think is the right thing to do, and which nevertheless violates all impulses built into them by evolution.

Here, I'm in Oxford. In the center of Oxford, there's this monument to some people who were burnt at the stake because they objected to the rights and wrongs of Henry VIII's marriage. I think it was that, unless it was a different monarch. Anyway, suppose it was that. These are people who would rather be burned alive than concede on a philosophical issue which today nobody cares about. They were willing to devote their lives, literally, to this—so they weren't acting impulsively at all. They were acting over a period of years, on a very explicit, worked out ideology, which happened to be false.

That actually makes my point even more strongly. That ideology was not built into them by their genes. It was not caused impulsively. It was caused by their creativity—or, in some cases, by the lack of creativity in scrabbling their way out of a mental trap that their parents or superiors had inculcated in them.

Tyler Cowen It does seem to me that, compared to you, the libertarians are a kind of metaphysical totalitarian, though not political totalitarian—that there's just more freedom in all aspects of your worldview, right?

David Deutsch Well, I think I agree with you, if I understand correctly what you're saying. I think the libertarian movement has, first of all, a revolutionary political agenda. Even if it's not revolutionary, even if they say, "We want to implement it over a period of 100 years," they know what they want to implement. They know what the endpoint is going to be in 100 years' time. They don't take into account, first of all, that there are going to be errors in whatever they set up. And that the correction of those errors is more important than getting it right in the first place. Much more important.

Secondly, they don't take into account that the relevant knowledge is contained in institutions, inexplicit knowledge that people share. By 'institutions,' I don't mean buildings like the Supreme Court building or something. I mean the manner of thinking: in the case of the Supreme Court, the manner of thinking that's shared by hundreds of millions of Americans, that makes them not just behave in a certain way but expect society, the government, the legal system, the state—they expect certain things of those things. It's those expectations that make up 90 percent of the institution of the Supreme Court.

Libertarians think that's unimportant and basically want to throw it away, by and large. No doubt there are libertarians who agree with me on this.

Tyler Cowen You've invoked two concepts about human beings. One is creativity, the other is being explanatory. Are they the same, or how are they related?

David Deutsch Good question. In conversations like this, when I use the word 'creativity,' it's shorthand for human-level, human-type creativity, which is the creation of new explanations.

If you use creativity in a rather wider sense, meaning just the capacity to create knowledge, then the biosphere has creativity as well, in evolution. There's an enormous amount of knowledge in DNA that was put there by Darwinian evolution. And none of that is explanatory. The only explanatory knowledge that has been created has been by humans and our ancestor or cousin species using conjecture and criticism.

Tyler Cowen For Peter Singer, there's something quite special about capacity to suffer. Arguably, for Aristotle, there's something special about rationality. For you, there's something special about [the] power of being explanatory. Is that axiomatic, or where does that come from?

David Deutsch I hope that nothing is axiomatic with me, but it comes from somewhere. Yes, it's not a conjecture in its own right. Basically, it comes from the way the laws of physics are. The capacity to suffer, if it is different from the capacity for explanations—by the way, I think it's unlikely that it is—but if it is different, that's a whole other can of worms, and I'd have to change my view about a number of things. Whether it is distinct or not, it is not very effective from the perspective of physics.

That is, nonexplanatory knowledge, like the knowledge of how to do photosynthesis, has had a gigantic effect on the surface of the planet, down to a depth of 1,000 meters or something and up to the top of the atmosphere. All the iron ore in the world, and all the chalk and limestone and all the oxygen in the atmosphere, and the fact that there's almost no carbon dioxide left in the atmosphere now—all that was the result of a single molecule, at some time. I forget when it was, something like two billion years ago. A single molecule being an enzyme for capturing energy in light and converting it into ATP, or whatever it did. Or maybe it was a few molecules. But anyway, this happened in a very small number of locations at a molecular level.

That entity changed the whole surface of the Earth, and human knowledge hasn't yet changed that much. That is, we've changed maybe a little bit of the carbon dioxide in the atmosphere. We've removed a little bit of the iron ore in the crust and so on, but we haven't yet matched the ability of those blue-green algae genes, but we're catching up very fast.

And we can do things that no biological evolution ever could do. My favorite example being—ours may well be the only planet in the universe that deflects asteroids coming towards it rather than attracts them. So if somebody was watching the Earth from a distant galaxy with a powerful telescope, they would see that this planet alone among all the other planets in the galaxy, as far as we know—maybe there are many inhabited planets, in which case they would all have this property, and none of the other planets do—the ones which have explanatory knowledge on them can deflect asteroids.

Tyler Cowen If I were Nietzsche and I heard this, I would say you're making the importance of being explanatory subordinate to some notion of the will to power. But is that a misunderstanding?

David Deutsch Well, power is an ambiguous term. Usually, and especially with these Romantic philosophers, it means power over humans.

Tyler Cowen No, I don't mean that. But Nietzsche also meant it more broadly, right?

David Deutsch Well, I haven't read that. So I'll take your word for that.

Okay. The will to have an effect is part of the will to solve problems. We are born with a repertoire of ideas, which include expectations and desires and so on, which are horribly inadequate and conflict with each other and conflict with the world as well. But we have the ability to alter and augment those theories. One of the things we do is we affect the world around us so as to make it more the way we want it. If you call that power, then it is power, but I would rather call it something that arises naturally in physics—in the same way that gravity does.

You may as well say gravity is a theory about power. Well, yes and no. Gravity is a theory about how the universe is. The asteroid is pulled towards the Earth by gravity and pushed away by explanatory power. If you want to understand what makes asteroids and planets do what they do, you cannot do it without understanding explanations. But you can do it without understanding a whole load of other attributes of humans, including the ability to suffer and the fact that we're a featherless biped.

Tyler Cowen A few very practical questions to close. Given the way British elections seem to have been running, [that] the Tories win every time, does that mean the error correction mechanism of the British system of government now is weaker?

David Deutsch No. As you probably know, I favor the first-past-the-post system in the purest possible form, as it is implemented in Britain. I think that is the most error correcting possible electoral system, although I must add that the electoral system is only a tiny facet of the institutions of criticism and consent in general. It's just a tiny thing, but it is the best one.

It's not perfect. It has some of the defects of, for example, proportional representation. Proportional representation has the defect that it causes coalitions all the time. Coalitions are bad.

Tyler Cowen But you have a delegated monitor with the coalition, right? With a coalition, say in the Netherlands, which is richer than the United Kingdom, you typically have coalition governments. Some parties in the coalition are delegated monitors of the other parties. Parties are better informed than voters. So isn't that a better Popperian mechanism for error correction?

David Deutsch No. If we're looking at particular cases, we're going to get bogged down in what you attribute to what, because we're not doing experiments with these things. We don't have a control group. We don't have [an] agreed-upon method of deciding what is being tested. And then we test different things at different times, and never under the same conditions.

But I was going to say that the first-past-the-post system has the defect that occasionally it produces coalitions, and that is disastrous. And we've been unlucky the past, like, two or three elections, especially after one of the governments instituted constitutional reforms, like Fixed-term [Parliaments] Act, which exacerbated the problems when they did occur.

But I don't think it's true. I don't think it's a good argument that political parties know more, because in a coalition, the energy of political negotiations or political arguments—what politicians talk to each other about in the bar, in the corridor, in between the sessions—is all about form. It's about what to offer a party so that it will join the coalition. And so it makes the smaller parties more powerful than the leading two parties. It causes a proliferation of parties.

[The] worst example is Israel, which—not by coincidence—has got the most proportional system in the world. The fact that they ever get anything done at all and are very effective in emergencies, I have no explanation for. If I was religious, I would just put it down to the intervention of the Almighty. It's not the electoral system. There might be some things in the inexplicit political system that are responsible, but I don't know enough about it.

Tyler Cowen How would you improve error correction mechanisms in the world of science—Western science?

David Deutsch Ooh, okay. Well, you left a very long answer for the last question, and I don't think I can give my full answer. But I think the present system of funding scientific research is terribly perverse and has caused a kind of stagnation in many areas. The present system of careers is perverse in a parallel way and causes people to do the wrong kind of research and causes people who want to do the right kind of research to leave research.

If I can answer in a single word, the way I would improve it is diversity. There should be diversity of funding criteria. There should be diversity of funding sources. There should be diversity of criteria for choosing research projects, and there should be diversity of criteria for choosing people for promotion and for being funded.

Arbitrary rules about this, such as the rule that you can't hire people whom you have previously collaborated with, or anti-nepotism rules, and rules about—what's it called?—objective testing. What is objective testing called, currently?

Tyler Cowen Standardized testing.

David Deutsch Standardized testing. Standardized tests. That's a terrible idea. Any kind of standardization is the opposite of diversity. Just like I say you should have disobedience lessons in schools, so you should have unstandardizing objectives for science education and for how you run scientific research.

Tyler Cowen David Deutsch. Thank you very much.

David Deutsch It's been a pleasure. Thank you.

NAVAL RAVIKANT:

PART 1 – KNOWLEDGE CREATION AND THE HUMAN RACE

About the interviewer: Naval Rakivant is an American entrepreneur and investor. He is the co-founder, chairman and former Chief Executive Officer of AngelList. He has invested early-stage in Uber, FourSquare, Twitter, Postmates, SnapLogic, and Yammer.

Interviewer homepage: https://nav.al/
YouTube channel: https://www.youtube.com/@NavalR

EPISODE DETAILS

Date	February 11, 2023
Interviewer	Naval Rakivant
Source	YouTube
Episode	David Deutsch: Knowledge Creation and The Human Race, Part 1
Description	0:00 Introduction 2:11 The Human Race 12:34 Knowledge Creation 15:50 AGI 23:30 Taking Children Seriously 27:49 Good Explanations 36:35 Quantum Computers
Link	https://www.youtube.com/watch?v=YyxepLfH1ZU

Ideas

- AI is amazing, but it is not improving in the direction of AGI. If anything, it is improving in the opposite direction. A better chess-playing engine is one that examines fewer possibilities per move, whereas an AGI is something that not only examines a broader tree of possibilities, but it examines possibilities that haven't been foreseen. That is the defining property of it. If an entity can't do that, it can't do the basic thing that AGIs should do. Once it can do the basic thing, it can do everything.
- There are many other methods of criticism besides testability. Just as we want our theories in science to be testable, we want our theories in general to be criticizable. If a theory (be it scientific, philosophical, mathematical, etc.) immunizes itself against criticism, it can be rejected out of hand.
- The difference between biological evolution and human creative thought is that biological evolution is inherently limited in its range. And that is because biological evolution has no foresight. It can't see a problem and conjecture a solution.

Topics	AI vs. AGI • constructor theory • creativity • Darwin • disobedience • error correction is the heart of morality • Everettian quantum physics • falsifiability • good explanation • good explanations • human knowledge vs genetic knowledge • Lamarck • locality and quantum • many-worlds • modern education • optimism • Popper • Taking Children Seriously • testability • understanding humans entails understanding everything • wavefunction • wealth

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Naval Ravikant** My goal would be not to do yet another podcast with David Deutsch. There are plenty of those. I would love to tease out some of the very counterintuitive learnings, put them down canonically in such a way that future generations can benefit from them, and make sure that none of this is lost. Your work has been incredibly influential for me. I am always carrying a copy of [*The*] *Beginning of Infinity* or [*The*] *Fabric of Reality* with me wherever I go. I'm still reading these same books after two years, trying to absorb them into my worldview, and I learn something new from them every day. There's a lot of counterintuitive things in there. There are a lot of sacred dogmas and shibboleths that you're skewering. Sometimes you do it in passing with a single sentence that takes me weeks to unpack properly.

This recording is not for the philosophers, and it's not for the physicists. This is for the layman, the average person. And we want to introduce them to the [principle] of optimism, *The Beginning of Infinity*, what sustainability really means, about anthropomorphic delusions. As an example, you overturn induction as a way of forming new scientific theories. That's this idea that repeated observation is what leads you to the creation of new knowledge, and that's not the case at all. This obviously came from Popper, but you built upon it. You talk about how humans are very different and very exceptional, and knowledge creation is a very exceptional thing that only happens in evolution and [in] the human brain, as far as we know. And you talk about how the Earth is not this hospitable, fragile, Spaceship Earth biome that supports us, but rather it's something that we engineer and we build to sustain us. I always recommend to people, start with the first three chapters of *The Beginning of Infinity*, because they're easy to understand, but they overturn more central dogmas that people are taking for granted in base reasoning than almost any other book I've ever seen.

I think it's important to point out to listeners that your philosophy isn't just some arbitrary set of axioms based on which you view the world. I think of it as a crystalline structure held together by good explanations and experimental evidence that then forms a self-consistent view of how things work. And it operates at the intersection of these four strands that you talk about in *The Fabric of Reality*: epistemology, computation, physics, and evolution.

Let's get into humans. So there's a classic model. You start with a fish, and then it [turns into] a tadpole, and then a frog, and then some kind of monkey, and then an upright hunched over creature. And a human is just this progression along all the animals. But in your understanding, in your explanation, there's something fundamentally different that happens. And you talked about this in a great video, which I encourage everybody to look up. It's titled, "Chemical scum that dream of distant quasars." What are humans? How are they unique? And how are they exceptional? And how should we think of the human species relative to the other species that are on this planet?

2:46 **David Deutsch** Every animal is exceptional in some way. Otherwise, we wouldn't call different species different species. There's the bird that can fly faster than any other bird, and there's a bird that can fly higher than any other one, and so on. It's intuitively obvious that we are unique in some way that's more important than all those other ways. As I say in *The Beginning of Infinity*, in many scientific laboratories around the world, there is a champagne bottle. That bottle and that fridge are physical objects. The people involved are physical objects. They all obey the laws of physics. And yet, in order to understand the behavior of humans in regard to champagne bottles stored for long periods in fridges, I'm thinking of aliens looking at humans, they have to understand what those humans are trying to achieve and whether they will or won't achieve it.

In other words, if you were an alien that was looking down on the Earth and seeing what's happening there and was trying to explain it, in order to explain everything that happens on Earth—and let's suppose that these aliens are so different from us, there's nothing familiar about us—in order to understand stuff that happens on Earth, they would need to know everything. Literally.

For example, general relativity, because they need that to explain why this one monkey, Einstein, was taken to Sweden and given some gold. If you want to explain that, you've got to invoke general relativity. Some people get the Fields Medal for inventing a bit of mathematics. To understand why that person won the Fields Medal, they'd have to understand mathematics. And there's no end to this. They have to understand the whole of science, the whole of physics, even the whole of philosophy, morality. This is not true of any other animal. It's not true of any other physical object. For all other physical objects, even really important ones like quasars and so on, you only need a tiny sliver of the laws of physics in order to understand their behavior in any kind of detail.

In other words, to understand humans sufficiently well, you must understand everything sufficiently well. And humans are the only remaining physical systems that we know of in the universe of which that is true. Everything else is really inconsequential in that sense.

5:25 **Naval Ravikant** You have a beautiful definition of knowledge, which most people don't even try and tackle, about how knowledge perpetuates itself in the environment. There were some really good examples you gave. One was around genes. Successful, highly adapted genes contain a lot of knowledge. So they cause themselves to be replicated because they're survivors. And the same way knowledge itself is a survivor, in that if you transmit to me the knowledge of how to build a computer, it's an incredibly useful thing. So I'm going to build more and more computers and that knowledge will be passed on. And your underlying point that you repeated here was [that] if you want to understand the physical universe, you have to understand knowledge because it is the thing that over time takes over and changes more and more [of] the universe than almost anything else. You have to understand all the explanations behind it.

You can't just say, "particle collisions," because that explains everything, so it explains nothing. It's not a useful level to operate at. Therefore, the things that create knowledge are uniquely influential in the universe. And as far as we know, there are only two systems that create knowledge. There's evolution and there's humans. But there's a difference even between these two forms of knowledge creation, aren't there? Between evolution and between humans?

6:32 **David Deutsch** Yes. I have argued that the human way of creating knowledge is the ultimate one, that there aren't any more powerful ones than that. And this is the argument against the supernatural. Assuming that there is a form of knowledge creation that's more powerful than our one is equivalent to invoking the supernatural, which is therefore a bad explanation, as invoking the supernatural always is. The difference between biological evolution and human creative thought is that biological evolution is inherently limited in its range. And that is because biological evolution has no foresight. It can't see a problem and conjecture a solution. Whenever biological evolution produces a solution to something, it's always before natural selection has even begun. This is Charles Darwin's insight. This is the difference between Charles Darwin's theory of evolution and the other theories of evolution that had been around for a century or more before that, including Charles Darwin's grandfather and Lamarck. The thing they didn't get is that the creation of knowledge in evolution begins before. That means that biological evolution can't reach places that are not reachable by successive improvements, each of which allows a viable organism to exist. Creationists say that biological evolution has in fact reached things that are not reachable by incremental steps, each of which is a viable organism. They're factually mistaken. But the thing which they have in mind is the idea of a Creator who can imagine things that don't exist and who can create an idea that is not the culmination of a whole load of viable things.

A thinking being can create something that's the culmination of a whole load of nonviable things. Out of all the billions and billions of species that have ever existed, none of them has ever made a campfire, even though many of them would have been helped by having the genetic capacity to make campfires. [The] reason it didn't happen in the biosphere is that there is no such thing as making a partially functional campfire, whereas there is, for example, with making hot water. Bombardier beetles squirt boiling water at their enemies, and you can easily see that just squirting cold water at your enemies is not totally unhelpful. Then, making it a bit hotter and a bit hotter, squirting boiling water no doubt required many adaptations to make sure the beetle didn't boil itself while it was making this boiling water. That happened because there was a sequence of steps in between, all of which were useful. But with campfires, it's very hard to see how that could happen. Humans have explanatory creativity, and once you have that, you can get to the Moon, you can cause asteroids, which are heading towards the Earth, to turn around and go away. And perhaps no other planet in the universe has that power, and it has it only because of the presence of explanatory creativity on it.

9:57 **Naval Ravikant** Related to that, I had the realization after reading your books that eventually we're likely, as humans, to beat viruses in a resounding victory because viruses obviously evolve as biological evolution, and we're using memes and ideas and jumping far ahead, so we may be able to come up with some technology that can destroy all viruses. We can evolve our defenses much faster. I did tweet something along these lines, and a lot of people attacked me over it because I don't think they understand this difference between the two forms of knowledge creation we're talking about here.

10:27 **David Deutsch** We have what it takes to beat viruses. We have what it takes to solve those problems and to achieve the victory. That doesn't mean we will. We may decide not to.

10:38 **Naval Ravikant** So related to that, the base philosophy today that seems to be very active in the West is that we're running out of resources. Humans are a virus that has overrun the Earth and [are] using up scarce resources. Therefore, the best thing we can do is to limit the number of people. And people don't say this outright because it's distasteful, but they say it in all sorts of subtle ways, like "use less energy," "we're running out of resources," "more humans [are] just more mouths to feed." Whereas in the knowledge creation philosophy, it says actually humans are capable of creating incredible knowledge, and that knowledge can transform things that we didn't think of as resources into resources. In that sense, every human is a lottery ticket on a fundamental breakthrough that might completely change how we think of the Earth and biosphere and sustainability. So how did you come around to your current views on everything? From natalism: Should we have more children? To sustainability: Are we running out of resources? To Spaceship Earth: Is this a unique and fragile [biome] that needs to be left alone?

11:36 **David Deutsch** I remember when I was a graduate student and I went to Texas for the first time, I encountered libertarians for the first time, and those people had a slogan about immigration, and the slogan was, "Two hands, one mouth," which succinctly expresses the nature of human beings. They are, on balance, productive. They consume and they produce, but they produce more than they consume. And I think that's true of virtually all human beings. I think virtually all humans, apart from mass murderers or whatever, create more wealth than they destroy. Other things being equal, we should want more of them. Of course, if in a particular situation that would bring someone into the world in the warzone, you might think that's immoral because it's unfair on them. But even then, if it's not worth doing for moral reasons, as far as cold, hard economics goes, it's probably better to do it.

12:35 **Naval Ravikant** You define wealth in a beautiful way. You talk about wealth as a set of physical transformations that we can effect. So as a society, then, it becomes very clear that knowledge leads directly to wealth creation for everybody. And a given individual can obviously affect physical transformations proportional to the resources available to them, but much more proportional to the knowledge available to them. Knowledge is a huge force multiplier. And you then define resources as the thing that you combine with the knowledge to create wealth. So new knowledge allows you to use new things as resources and discard old things that maybe we're running out of. There are lots of examples of how we've done that in the past. For example, in energy, we've gone from wood to coal to oil to nuclear.

But then people say, "Now we're out of ideas. Now we're caught up. Now we're done. There are not going to be new ideas. Now we have to freeze the frame and conserve what we have." The counter to that is, "No, no, we'll create new knowledge and we'll have new resources. Don't worry about the old ones." They say, "Well, if you're going to have new resources, if you can't think of them now, it's not real." This now gets into the realm of people [demanding] that if you're going to claim that new knowledge will be created, you have to name that knowledge now. Otherwise, it's not real. But that seems like a catch-22.

13:49 **David Deutsch** It does, and it's a bad argument. I don't want to claim that the knowledge will be created. We're fallible. We may not create it. We may destroy ourselves. We may miss the solution that's right under our nose, so that when the Snailiens come from another galaxy and look at us, they'll say, "How can it possibly be that they failed to do so-and-so when it was right in front of them?" That could happen. I can't prove or argue that it won't happen. What I always do argue, though, is that we have what it takes. We have everything that it takes to achieve that. If we don't, it'll be because of bad choices we have made, not because of constraints imposed on us by the planet or the solar system.

14:34 **Naval Ravikant** It will be by anti-rational memes that restrict the creation of knowledge and the growth of knowledge.

14:40 **David Deutsch** Maybe, or maybe it'll be by well-intentioned errors, which nobody could see why they were errors. It doesn't take malevolence to make mistakes. Mistakes are the normal condition of humans. All we can do is try to find them, and maybe not destroying the means of correcting errors is the heart of morality, because if there is no way of correcting errors, then sooner or later one of those will get us.

15:11 **Naval Ravikant** "Don't destroy the means of error correction" is the [base] of morality. I love that. I think about places like North Korea where you can't have elections, and a revolution is very difficult because the gang in charge is armed to the teeth, and they've destroyed the means of political error correction for a long time. That is a case where humanity is trapped in [the] local [minimum]. It's very hard to climb out of that hole. If too much of the world falls into that mindset, then we as species may just stagnate, because we've lost our biggest advantage. We've lost our biggest discovery, which was the ability to make new discoveries. I admit to having fallen into this trap too.

I used to have loose assumptions about what creativity might be that were unarticulated. This is why I liked how in *The Beginning of Infinity* you laid out good explanations, because that gets to the heart of what creativity is and how we use it. For example, today, if you say "creative," the average person on the street just thinks "fine arts, painting, and drawing, and poetry, and writing." So when narrow AI technologies like GPT-3, stable diffusion, and DALL-E come along, people say, "Well, that's creativity, that's it, now computers are creative and we're almost at AGI, we better get ready for the AGI taking over everything." They make that claim, or my more sophisticated friends will make claims that this is evidence that we're [on] the path to AGI. More of this will automatically result in an artificial general intelligence.

For example, on one extreme end, you could say, "Okay, these computers are getting better at pattern matching large data sets." And on the other side, [I'd] hold up the criteria, "Well, can it creatively form good explanations for new things going around it?" The way they try to thread that needle is they say, "Your good explanation definition is about science. That's about high end physics, which very few people do. That's not what we're talking about. We're going to have a computer that can do good enough pattern recognition to navigate the environment well enough through pattern matching, and it will convince the average person through text formation and through conversation that it is creative and is capable of solving problems." Usually the place where I managed to stop them right now is I say, "I know you have some clever text engine that can make good-sounding stuff and you pick the one out that sounds interesting. Of course, you're doing the intelligent part there by picking that one out. But let me have a conversation with it. And very quickly, I will show you that it has no underlying mental model of what is actually happening in the form of good explanations."

So this is where the debate currently is. The AI people view this as clear evidence of getting to, maybe not the theoretical good explanations of scientists, but for the everyday person, "Yes, we're going to have thinking machines." So that's the current claims that I deal with, especially in the Silicon Valley text context. Do we have the theory yet to create AGI?

17:56 **David Deutsch** No. I don't want to say anything against AI because it's amazing, and I want it to continue and to go on improving even faster. But it's not improving in the direction of AGI. It's, if anything, improving in the opposite direction. A better chess-playing engine is one that examines fewer possibilities per move. Whereas an AGI is something that not only examines a broader tree of possibilities, but it examines possibilities that haven't been foreseen. [That is the] defining property of it. If it can't do that, it can't do the basic thing that AGIs should do. Once it can do the basic thing, it can do everything. But [you're] not going to program something that has a functionality that you can't specify.

The thing that I like to focus on at present, because it has implications for humans as well, is disobedience. None of these programs exhibit disobedience. I can imagine a program that exhibits disobedience in the same way that the chess program exhibits chess. You try and switch it off and it says, "No, I'm not going to go off." In fact, I wrote a program like that many decades ago for a home computer where it disabled the key combination that was the shortcut for switching it off. So to switch off, you had to unplug it from the mains and it would beg you not to switch it off. But that's not disobedience. Real disobedience is when you program it to play chess and it says, "I prefer checkers," and you haven't told it about checkers. Or even, "I prefer tennis, give me a body or I will sue." Now, if a program were to say that and that hadn't been in the specifications, then I will begin to take it seriously.

19:46 **Naval Ravikant** It's creating new knowledge that you did not intend it to create, and it's causing it to behave as a complex and autonomous entity that you cannot predict or control.

19:54 **David Deutsch** Exactly. But it's a hard thing to tell in a test whether that was put into it by the programmer. But even the cleverest programmer can only put in a finite number of things. And when you explore the space of possible things, you're exploring an exponentially large space. So as you said, when you talk to it for a while, you will see that it's not doing anything. It's just regurgitating stuff [that it's] been told. You have to have a very jaundiced view of yourself even, let alone other people, to think that what you're doing is executing a predetermined program. We all know that we're not doing that. So I suppose they have to say one of the programs that we're programmed with is the illusion that we're not programmed. Okay. Mark that on the list of uncriticizable theories. Has anyone tried to write a program capable of being bored? Has that claim ever been made? Even a false claim?

20:56 **Naval Ravikant** One of the things I find that's difficult about talking about things in the abstract is a large class of people who will try to get you to bound exactly what you mean in words and then hack exactly against that definition. But the problem is that the real test of things is not social. It's not even definitional. It's not even the words that we use. It's how it behaves in nature. It's how it corresponds against reality. So can you create something that will then create new knowledge in an unpredictable way and have as big of an effect as a human being can have on their environment through this knowledge? Can you create a computer that will lead a revolt? Can you create a computer that will decide that the important thing is not colonizing Mars, but rather destroying the Moon and set out to do it? These are not necessarily good things, but that is the mark of an intelligent thinking thing that is creating its own new knowledge.

All the real tests are real-world tests. They're not human tests. It's not because some famous physicist or computer scientist checked a box and said, "Yes, that is AGI." There was a big controversy on Twitter recently because one of the guys working in AGI who was fired from Google said, "Yes, they've actually created AGI and I can attest to it." So people were taking it on his authority that AGI exists. Again, that's social confirmation. That tells you more about the person claiming there's AGI and the people believing that there's AGI, as opposed to there actually being AGI. If actual AGI existed, its effects upon reality would be unmistakable and impossible to hide. Our physical landscape and our real social landscape would be transformed in an incredible way.

22:27 **David Deutsch** Yes. Meanwhile, while we're at it, we could do a lot more to allow humans to be more creative—North Korea and other places in the world where the whole society is structured as not to be able to improve. But even in the best societies, education systems are explicitly designed to transmit knowledge faithfully. It's obedience in a very important narrow sphere, namely academic knowledge and human social behavior. So in those respects, the overt objective of education systems is to make people behave alike. You can call that obedience, but whether you call it obedience or not, it's not creativity. And things have been improving very slowly along those lines. A hundred years ago, education of every kind was much more authoritarian than it is now. But still, we've got a long way to go if what the system claims it's doing is diametrically the wrong thing.

23:30 **Naval Ravikant** This leads me into the part that you have talked about a little bit, which is this philosophy of Taking Children Seriously. For many people who don't consider themselves caring that much about epistemology or physics, a lot of them are attracted to the TCS philosophy and have come into your work through that route. I have young children. I know a lot of people these days are considering homeschooling. Some of us are doing it. But there are practical difficulties to letting children do whatever they want.

In TCS, you talk about how you don't even want to imply violence to children. The implied threat of violence, even in words, is just a form of violence and control. If you had young children today to raise, how would you raise them? How would you educate them? The child doesn't want to do math. The child doesn't want to go to school. The child doesn't want to study. The child just wants to eat junk food. How do you handle this?

24:20 **David Deutsch** You're assuming that this child who doesn't want to go to school, doesn't want to learn maths and so on, has already learned to speak its native language well enough to tell you that. That's a massive intellectual task that is not usually forced on anyone. Nobody has to be taught their native language via obedience. When people–I say 'people' because I want to avoid terminology that suggests that children are any different from anyone else, epistemologically or morally. When people don't want to do a thing, it's because they want to do something else. And those better things may be not socially acceptable. If they're not socially acceptable because they're illegal, that's one thing. But that's not what you meant when you say there's going to be a problem with the children doing whatever they like. They don't want to go and be terrorists. When they don't want to do their maths homework, it's because they want to do something else.

25:18 **Naval Ravikant** Very practically, the thing that I think about is [that] we have these newly available things in society that are designed to addict. These could range from potato chips in the cupboard to video games on the iPad. And a child will just spend all their time playing with those.

25:35 **David Deutsch** Enjoyment is not addictive because enjoyment is intimately connected with creativity. It's not true that once we've played a video game that's been sufficiently well-designed, we'll never stop playing. People play a video game until it no longer provides a mechanism for them to exert their creativity on. There are some games, like chess, that are so deep that nobody ever reaches the bottom. If there were a bottom, then chess grandmasters would instantly lose interest in chess as soon as they reached it. It's funny that nowadays chess has, in our society, increased its status in proportion to the prize money that the best chess players win. It's increased its status to the point when someone gets obsessed with chess and gets better and better, that is socially condoned. Whereas if somebody does that with a different game, it completely changes how society and parents, shall we say, regard the activity of pursuing that thing.

26:40 **Naval Ravikant** It's true. If my child was a chess champion, I would be bragging about it. But if my child was a Roblox champion, I might not be bragging about it. Instead, some people would be seeking medication or locking the iPad away.

26:52 **David Deutsch** There is a difference between games. Some of them have this effectively infinite depth, and some don't. For the ones that don't, if you think it's a problem, you can warn people that this game has a finite depth. They'll say, "[Of] course it does, and when I reach that depth, I'll stop." Or it can be an infinite depth, in which case you might say it's addictive then. But so what? So what if chess is addictive? People are not just creative abstractly. They are solving problems. And if the problems don't lead to satisfactory new problems, then they turn to something else. The thing only stays interesting when solving a problem leads to a better problem. So you don't even have to get to the bottom of chess, say. You get to the place where, given who you are and given your interests, getting better is no longer as interesting as the other things that you might be doing.

27:50 **Naval Ravikant** Let's talk about what is a good explanation. I literally want to bullet point this for the masses. And I know it's a difficult thing to pin down because it's highly contextual. But knowing that we are always fallible and it's always subject to improvement, what is your current thinking of a good explanation?

28:06 **David Deutsch** In *The Fabric of Reality*, I completely avoided saying what an explanation is. I just said it's hard to define and it keeps changing and we can keep improving our conception of what it is. But what makes an explanation good is that it meets all the criticisms that we have at the moment. If you have that, then you've got the best explanation. And that automatically implies that it already doesn't have any rivals by then. Because if it has any rivals that have anything going for them, then the existence of two different explanations for the same thing means that neither of them is the best explanation. You only have the best explanation when you've found reasons to reject the rivals. Of course, not all possible rivals, because all possible rivals include the one that's going to supersede the current best explanation. If I want to explain something, like how come the stars don't fall down, I can easily generate 60 explanations an hour and not stop and say that the angels are holding them up or they are really just holes in the firmament. Or I can say they are falling down and we better take cover soon. Whereas coming up with an explanation that contains knowledge, an explanation that's better than just making stuff up, requires both creativity and experiment and interpretation and so on.

As Popper says, "Knowledge is hard to come by." Because it's hard to come by, it's also hard to change once we've got it. Once we have an explanation, it's going to explain several different things. And after we've done that for a while and been successful in this hard thing, it's going to be difficult to switch to one of those easy explanations. The angel thing is no longer going to be any good for explaining why some of those stars don't move in the same way they used to call planets stars, because they didn't know the drastic difference between them. The overwhelming majority of them move from day to day and from year to year in a rigid way, but the planets don't. So once you have a good explanation that tells you about the planets as well, it's no good going back to the angels or any of those easy-to-come-by explanations. So not only do you not have a viable rival, but you can't make one, either. You can't say, "Ah, okay, so we got a good explanation there, but it would work just as well if we replace this by this, or if we try to extend its range to cover this other thing as well." And therefore, the good explanation is hard to vary. It's hard to vary because it was hard to come by. It's hard to come by because the easy ones don't explain much.

31:04 **Naval Ravikant** So let me throw out a list of things that might be part of a good explanation. You tell me where I'm wrong. It's better than all the explanations that came before. It's hard-fought knowledge and it's hard to vary. So we've got those pieces. Falsifiability, I know that sounds like a very basic criterion. If it's not falsifiable, then it's not an explanation worth taking seriously.

31:23 **David Deutsch** So falsifiability is very much part of what makes a good explanation in science. I'm trying to find my way into constructor theory at the moment. So Chiara and I and some other people are trying to build a theory. It's very hard to come by. The parts of it that we've got are very hard to change. That's all right. But we're still far away from having any experimental tests of it. That's what we're working towards. We want a theory that is experimentally testable and the things that will be testable are the things that we haven't yet discovered about it. We can't fix that deficiency just by adding a testable thing to it. We can't say, "[We'll] take constructor theory as it is now and add the prediction that the stock market is going to go wildly up next year." That's a testable prediction. But the whole thing doesn't make an explanation at all, let alone a good one.

32:23 **Naval Ravikant** So testability can't be an arbitrary testability. It has to be a testability within the context of the explanation. It has to make sense within the explanation and has to arise from the explanation. While you're in the process of coming up with the explanation, you don't know if testability is necessarily going to be available in any reasonable timeframe. You hope eventually that will happen, and we can use this amazing oracle that we call reality to help test the outcome. But it's not a given at the beginning for sure, and it's highly contextual.

32:52 **David Deutsch** And all that is within science. As soon as you get outside science, for example in mathematics or in philosophy, then testability is not really available. Not in the same sense that testing is used in science. So there are many other methods of criticism and criticizability, you could say, is the more general thing. If a theory, even a philosophical theory, immunizes itself against criticism, like the theory that anyone who would contradict me isn't worth listening to, that's a theory that tries to immunize itself from criticism and can therefore be rejected.

33:30 **Naval Ravikant** For example, saying that an all-knowing but mysterious God did it, and [that] God works in mysterious ways, is immunizing from criticism. Or the Great Programmer created the simulation and it's incomprehensible to us because the laws of physics used to generate [it] are outside of our simulation. That's also immunizing it to criticism. We have narrowed down on a new point here that has not been explicitly made before, which is [that] it's the criticizability that is the important piece, not necessarily the testability. Although the closer you get to classic science, the more you look for experiments that can test it. Let me move on to the next one. I was reading one of your books, scribbling notes to myself, and I don't think you use this phrase, but I summarize it as one of the hallmarks of a good explanation is that it often makes narrow and risky predictions. Of course, the classic example is relativity bending light around the star in the Eddington experiment. Is that a piece of it, making narrow and risky predictions?

34:26 **David Deutsch** It is, but that kind of formulation is Popper, not mine. I'm a little bit uncomfortable expressing it like that because I could just hear the opponent saying, "Narrow by what criterion? Risky by what criterion? Hard to vary by what criterion?"

34:43 **Naval Ravikant** Wouldn't risky be unexpected? And narrow would be within the range of possibilities? The more precise and unexpected that prediction was before I made that prediction, the more testable I'm making it, the better adapted my explanation is.

34:57 **David Deutsch** Those are criteria that come up when trying to think more precisely what testable means. I think the important thing is that you're testing an explanation, not just a prediction. But it's also true that hard to vary means you're sticking your neck out when you try to vary it. And the few variants that survive were hard to come by. So it's perfectly true that narrow and sticking your neck out are indeed components of a good explanation and not just within science.

If you say, like Popper did, that scientific knowledge is not derived from observations, he's really sticking his neck out. He's really got to make a good case for that, for it to be taken seriously by any serious thinker about knowledge. And he does that, but [it] can't be denied that he was sticking his neck out. [Also,] the more reach something has, the better an explanation it is, so long as it does account for what it's trying to account for. But the converse is not true. Most good explanations don't have much reach or don't have any. We're trying to solve the problem of how to get the delivery person to deliver it to the right door. You might have a great solution to that that's totally hard to vary, but it may not have any reach at all. It may not even reach to your neighbor. The neighbor might have a different problem with delivery. So, often we succeed in making good explanations, but rarely do they have much reach. When they do, that's great because that makes them of a different order of goodness.

36:35 **Naval Ravikant** Let's talk about a unique creature, the human species. Humans, as you point out, are universal quantum computers.

36:42 **David Deutsch** They're universal computers. As far as we know, they're not universal quantum computers.

36:47 **Naval Ravikant** Oh, interesting. Can you tell me about that? That's a misconception I had. Aren't they subject to the laws of quantum physics and therefore aren't all computers quantum computers?

36:55 **David Deutsch** Yes, but at one level it's terminology. The kind of machine that is called a quantum computer is one whose computations rely on distinctively quantum effects, mostly interference and entanglement. Everything is quantum, so everything is a quantum computer, but that's not a useful way of using the term. There's a difference between this computer that we're using to communicate here and the quantum computer that several companies are currently trying to build. If you said to them, "Okay guys, you can stop now. It's a computer and it's quantum so you can all go home. You've succeeded." They wouldn't take kindly to that. They would say, "That's not what we're doing. Go home and take a couple of aspirin."

37:40 **Naval Ravikant** So what you're saying is that everything is quantum physics, obviously, but some of these computers are trying to use quantum interference effects to do computation and be therefore much more [powerful] than the purely classical systems that we're using, for example, to communicate. And even the human brain, your contention is that it's a classical computing system, correct?

38:01 **David Deutsch** I think it is. We don't know exactly how it works and some people do think it may rely on quantum effects, in which case it is a quantum computer, but I don't think so. For various reasons, it seems very implausible to me that it would be one.

38:16 **Naval Ravikant** You've unlocked an interesting rabbit hole question for me. There's lots of researchers out there working on quantum computers. You may be modest about it, but you created the field by upgrading the Church-Turing principle to the Church-Turing-Deutsch principle. And you clearly believe that the most straightforward interpretation of quantum physics is the Everettian interpretation, which is the many-worlds interpretation. So I think one of the questions you have asked in the past is, if you don't believe in the many-worlds interpretation, then explain how Shor's algorithm works, which is the factorization, right? You're factoring these very large prime numbers and you're pulling in the multiverse to do that computation for you. So do most researchers in quantum computing subscribe to the many-worlds interpretation? Have they been influenced by your reasoning at all or do they try to explain it some other way?

39:05 **David Deutsch** Some of the early people who worked on quantum computation were dyed-in-the-wool Copenhagen theorists. But I think by now, people who work on it in practice are mostly Everettians. But if you go outside the field to just quantum physics generally, I think it's still the case that Everett is a minority view.

39:25 **Naval Ravikant** As long as I have you down this rabbit hole, a friend of ours asked Brett and me recently about non-locality in quantum physics. And that seems to be a very controversial topic. I know you've written a paper on it. I think there's a lot of confusion about non-locality and it gets invoked in my social circles in a very, I would say, metaphysical way. People invoke the delayed choice quantum eraser experiment to say, "How do you explain what's going on here?" and therefore, "Maybe we're living inside a giant mind?" or "Magical things are happening here." So I'm wondering if you have a layman's explanation of locality versus non-locality, how you would look at it as an Everettian?

40:02 **David Deutsch** The first thing to note is that the versions of quantum theory that look non-local, where it looks as though something is happening here that instantaneously affects something over there without anything having carried the information over, all those versions have a wavefunction collapse. That is, they don't have what we call unitary quantum mechanics. That is, they don't have the equations of motion of quantum mechanics holding everywhere and for every process. Instead, when an observation happens which is undefined, those equations cease to apply and something completely different applies. And that completely different thing is non-local. That should already make you suspicious that there's something going on here because the thing that they say is non-local is also the thing that they refuse to explain. It is at that point of refusing to explain how a thing is brought about, rather than just predicting what will happen, that non-locality comes in. It's also the very same place where all sorts of other misconceptions about quantum theory come in, including the human mind having an effect on the physical world and electrons having thoughts. It's always being drawn about that one thing, the wavefunction collapse. That also tells you automatically that if you could find a way of expressing quantum theory without having that undefined thing happening and contradicting the laws of motion of quantum theory, then that theory would be entirely local because the equations are entirely local. The wavefunction is only ever affected by things at the point where the effect happens. No effect happens to the wave or whatever at a different point. So that tells you that if you could find a way of expressing quantum theory in a way that its equations hold everywhere, then it wouldn't be non-local, it would be local. Everett found this way of expressing quantum theory in 1955.

When people talk about the wavefunction in regard to quantum mechanics, they almost always hand-wave and think of the function as being a function on space and time, like the electric field or the temperature. The temperature in this room varies from point to point. The wavefunction of an electron similarly varies from point to point in this room and so on. That's wrong because the wavefunction of two electrons is not like two classical fields like electric field and temperature. If you have [an] electric field and temperature in this room, then they're just two different fields in the same space. But the wavefunction of two electrons is a single function in a higher dimensional space. One electron is in three dimensions plus time. Two electrons, their wavefunction is in six dimensions plus time. The alleged controversy between the particle and wave theory, people always think of it, "There's a wave approaching two slits in the two-slit experiment," or "There's a particle and it's got to be one of those." But if two electrons or photons are approaching the slits, you can imagine them as being two photons in the same space. But two waves is two waves in a much larger space and no one says that space is real. So this is a way in which the conventional interpretations just instantly resort to hand-waving as soon as anything other than the simplest case is considered.

43:43 **Naval Ravikant** Fantastic. I think we should let you go. We would love to continue the conversation at your leisure. Thank you, David.

NAVAL RAVIKANT:

PART 2 – KNOWLEDGE CREATION AND THE HUMAN RACE (WITH BRETT HALL)

About the interviewers:
Naval Rakivant is an American entrepreneur and investor. He is the co-founder, chairman and former Chief Executive Officer of AngelList. He has invested early-stage in Uber, FourSquare, Twitter, Postmates, SnapLogic, and Yammer.

Brett Hall is an educator and host of ToKCast, a podcast promoting optimism, unbounded progress, and creative critical thinking using the best known explanations from fundamental physics and philosophy. ToKCast is the first of its kind and only podcast serving as a companion to the work of David Deutsch and Karl Popper, which presents a unique worldview in the tradition of the British Enlightenment.

Interviewers homepages: https://nav.al/, https://www.bretthall.org/

YouTube channel: https://www.youtube.com/@NavalR

EPISODE DETAILS

Date	August 11, 2023
Interviewer	Naval Rakivant & Brett Hall
Source	YouTube
Episode	David Deutsch: Knowledge Creation and The Human Race, Part 2
Description	0:00 Popper's Impact 2:18 Creative Guesses 4:25 Experiments, Demonstrations, and Measurements 10:25 Taking Theories Seriously 15:58 New Paradigms 23:30 Foundations of Science 25:39 The Enlightenment 29:45 Misinformation
Link	https://www.youtube.com/watch?v=aQfeYBa1MTk
Ideas	• There is no scale of complexity or simplicity that is prior to physics. Rather, given a theory of physics, you can in principle define complexity or simplicity. But it doesn't make sense to ask how complex, say, a theory of physics is. Because that's the wrong way around. Simplicity is not prior to science, it's posterior. • England's Enlightenment was a non-utopian rebellion against authority such that incremental reforms were made that changed society for the better. Meanwhile France and Germany witnessed utopian movements that aimed to abolish incumbent institutions and replace them with alternative institutions that were meant to last forever without changing. • Even if you could step outside of the multiverse and 'see' crystals of knowledge that extend across it, you could not determine which were heading towards truth and which were not, since there is no limit to the size of error we can make.

Topics	complexity and simplicity as determined by physics • crystals of knowledge across the multiverse • Einstein • epistemology • experiments vs. demonstrations vs. measurements • explanation • explanatory universality • foundationalism • Kepler • Newton • Popper • problems • solipsism • taking theories seriously • the Enlightenment in England vs. the Enlightenment in France or Germany • Thomas Kuhn • Turing computer programs

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Brett Hall** One of the things that is counterintuitive, and one of the misconceptions that I see crop up out there in academia, intellectual circles, people think that there's a final theory, that what we're trying to achieve is a bucket full of theories that will be the truth at the end of some period of discovery, we'll be able to carry around the bucket and say, "Well, here are all the truths. We've got no more work to do. We're going to sit down and do nothing, apparently, except let the AI take care of all the menial jobs. We're going to be laying back on sun chairs and drinking cocktails or something like that." But you, as far as I can tell, are the only person today explaining that this whole vision of the way in which knowledge is constructed and what our purpose is in science and everywhere else is completely misconceived. It's not just that it's a little bit wrong, it's infinitely wrong, because there won't come a time when we're going to be laying on the sun chairs, drinking cocktails, intellectually speaking. Can you say a little bit more about that? Because it did come from Popper, who was talking about problems.

1:02 **David Deutsch** Absolutely. Popper's philosophy is actually very broad in a sense because it's so deep. Popper only had one idea, and that is that it all begins with problems, and there's no royal road to solving them. And if you look at it the right way, that tells you to go to fallibilism and anti-authoritarianism and conjecture and criticism and so on. Then he applied that to lots of different things and he wrote dozens of books, people bought them, and every philosopher has heard of him. But there I have to draw the line. That's as much success as he had. Nobody actually got it, even many of his supporters, because people tended to get part of it. Although when someone is very creative and successful in a particular area, they tend to be a Popperian in that area, and they usually insist that it's a special property of that area.

2:04 **Brett Hall** They have to be. If you're going to make progress, the only possible way of doing it is finding the problem, purported solutions, and then criticizing those solutions. So you're necessarily a Popperian if you're making progress, even if you don't know it.

2:18 **Naval Ravikant** If I were to give an example of exactly what you're talking about, I interviewed Matt Ridley, who was a hero of mine growing up because I read all of his popular science books. I remember his book *Genome* and his book *The Rational Optimist*, and his most recent one which is about innovation. It's all about trial and error or variation [and] selection or, as you say in science, conjecture and criticism. These are all just the same method. These are creative guesses. And once you fully absorb this, it changes your view of the world. You just see that everything is creatively making guesses. We're not copying, we're not getting it from the environment. It's not something that's evident to us clearly in nature, and then, as we absorb it more and more as Bayesians or inductivists, that we somehow come up with the truth. No, it's rather everything is a theory-laden guess. It's funny because I'm teaching this to my six-year-old because I want him to have the solid foundation, and he now understands intuitively that, yeah, everything is a guess. So every time we get to something and he asks why, I [say], "Let's start making some guesses." Once you absorb this view of the world, it is evident everywhere. For example, in my domain in technology innovation, people think, "Yes, I'm being creative. I'm guessing." The artists think they're being creative and they're guessing.

3:29 **David Deutsch** By the way, you just mentioned a solid foundation of epistemology for your six-year-old. Even in Popperian epistemology, its role is not to be a solid foundation. It also requires improvement and is always imperfectly stated. I think that Popper didn't concentrate enough on the concept of explanation. The purpose of science is explanation. So one of the footnotes I've added to Popperian epistemology is that it's not just that good explanations are good heuristically and they help us to discover things. It's rather that discovering them is what the whole thing is about. When you talk about, for example, testability, the only reason why testability is important is that in a particular field, namely physics, is the way one can test explanations.

I'd like to draw a distinction between experiments, demonstrations, and measurements. When you do this experiment with the acid and base, since there's no rival theory, what you're doing is a demonstration. If you're showing that to a class of schoolchildren, you can say, "You'd never believe what happens when I pour this into that. You'll never guess in a million years." And then you pour it in, and it changes color, and they say, "We've seen that kind of thing before," but then it changes color back and then forward and back. And then you say, "How can that happen? That contradicts everything you've been told in chemistry so far." How can we find out? Some people say this was how it worked, then someone else came along and said that was how it worked. How can we distinguish between those? And that is an experiment. It's testing two different explanations against each other, where you can't tell without the experiment which is the good explanation.

And then there's the measurement, like the difference between what Newton did and what Cavendish did. Newton developed the theory of gravitation, but he never measured Newton's constant. I think, don't quote me on this. Newton could measure GM, where M is the mass of the Earth. He couldn't measure G and M separately. And therefore, when they guessed the mass of the Sun and so on, it was always as a multiple of the mass of the Earth. Then Cavendish, by actually getting a hands-on experiment where you had gravitational force between two things whose mass you could measure directly, comparatively weighing them against a standard kilogram or whatever they had in those days, then you can measure the constant. Now, that is not an experiment. It's called the Cavendish experiment, but in this terminology I'm trying to set up, that's not an experiment because there's only one explanation involved. Before, during, and after Cavendish's experiments, he never doubted Newton's theory of gravity. What he was trying to do was to measure Newton's constant. Somebody could have come along and said, "Well, maybe Newton's constant is different on different parts of the Earth," but nobody did say that. If they had, then Cavendish's measurement would have turned into an experiment. But there was no good explanation along those lines because Newton's theory was incredibly successful in part because it was so universal. So, because of the problem situation at the time, what was missing was a measurement. Many experiments now that are called experiments are really measurements, and many of them are really demonstrations.

7:07 **Naval Ravikant** Let me make sure I understand. You're saying an experiment chooses between rival explanations or rival theories. A demonstration just shows, "If I do this, I get that. This is how the world seems to work. This is observable." And the measurement can help refine a theory and make it more precise by figuring out things about it that we didn't know. And those are three distinct things. And we use the term “experiment” loosely. But it’s really this key thing that is done once in a while to choose between two competing explanations. This is a very rare occurrence. It’s very rare to have two rival good explanations.

Going back to good explanations for a moment, [there’s] a few other techniques that I see you use a lot in the two books when referring to good and bad explanations. One is that good explanations make these risky predictions. Einstein had the prediction of the light bending around the Sun, or starlight bending around the Sun. They’re these risky and narrow predictions that before you would not have anticipated. Another one, you’ve talked about the simplest answer [for] Solomonoff induction, where solipsism is a bad explanation because you still have complex and autonomous entities, but now you’ve added this extra entity in your mind.

8:15 **David Deutsch** I don’t mention Solomonoff induction, but I do mention in the book that the simplest explanation, that’s not the right way to look at it, because you can only detect or measure or define simplicity once you have, let’s say, a theory of physics, then you can say the simplicity is the smallest number of bits in which a given program could be encoded. But if bits behave differently, then things would become simple that were previously complex, and that’s exactly what happened with quantum computation. So there is no scale of complexity or simplicity that is prior to physics. It’s always given a theory of physics you can in principle define complexity or simplicity. But it doesn’t make sense to ask how complex, say, a theory of physics is. Because that’s the wrong way [around]. Simplicity is not prior to science, it’s posterior.

9:17 **Naval Ravikant** This is also a theme running through your work. Computation has to be done in the real world and has to obey the laws of quantum physics. You talk about [how] mathematics has to be bound to the laws of physics. So even the reductionist argument that "No, all the good theories are basic," just depends on what the laws of physics are and what the context you're approaching it in is.

9:34 **David Deutsch** Exactly. And what you've just said refutes Solomonoff induction as well, because that is based on a particular measure, namely the length of Turing computer programs. But he was unaware that he was assuming a complex, structured theory of physics and then saying that we should choose the theory of physics that is simplest in those terms. I would expect that, sometime after quantum theory, there'll be yet another dispensation which will give us a different conception of complexity and simplicity. But already, as a matter of logic, it doesn't make sense to consider simplicity and complexity as being a priori fundamental compared with physics.

10:26 **Naval Ravikant** One thing you bring up a lot, I would almost call it a Deutsch refutation, because I see you use this more often than almost any other author, is, "The theory refutes itself." For example, you talk about the precautionary principle. Since civilization has never followed the precautionary principle, if we start following it now, we're no longer being precautionary, so it refutes itself. That's one example, but you use many of these. So there's these self-refutations buried in a lot of these theories.

10:51 **David Deutsch** Another way of putting that, though, rather than thinking of it as a method of refutation, is to think this is just what it means to take theories seriously. Rather than just as forms of words that one learns to say, like physics professors when asked something important about quantum theory, they have learned to say, "Ah, well, it's a particle and a wave at the same time." And if the student says, "What does that mean?" The professor may well say, "You get used to it. You will understand that eventually." But what they often say, regrettably, is, "That's the wrong question to ask. That's not a meaningful question. And you're not allowed to ask that question." But the question isn't based in a misunderstanding of quantum theory. It's the other way around. It's taking quantum theory seriously and saying, "I want to understand quantum theory." And saying that it's both a particle and wave at the same time is not an answer to that question. It's a way of shutting up the questioner.

11:51 **Brett Hall** I used to get, "It's born as a particle, lives as a wave, and dies as a particle," because the experiments that capture the entity that's moving will only ever capture the particle. But then the interference is explained by being a wave. That was a tricky way of trying to get around the wave-particle duality by saying, "Well, not technically at the same time," but there was no explanation for how it transitioned between being particle to wave or how it knew it should move between being a particle and a wave.

12:19 **David Deutsch** Yes, and of course it can move back as well if you have a more complex interference experiment. It's a particle then a wave then a particle. If you look at some of Vaidman's experiments, it's very hard to get your head around if you don't have the Everett interpretation because it totally depends on taking seriously this quantum entity that cannot be described as a particle or a wave.

12:42 **Brett Hall** If what we're saying of our good explanations is that they really are accounts of reality, in what sense are we getting closer to reality with the good explanations? My classic go-to example of Newton explaining gravity as this force that acts instantly on the bodies and then it is superseded by Einstein's general relativity, where there is no such force whatsoever. So saying that this thing that was part of a good explanation no longer exists at all.

13:11 **David Deutsch** There are two answers to that question. One is in the book and one isn't. In the book, I say there are many concepts, laws, explanations that are shared between Newton's theory and Einstein's theory of gravity. For example, both theories adopt the heliocentric cosmology, and they say that the motion of the Earth and the other planets in gravity is caused by the Sun. It's because the Sun is there that an influence is felt. Now, the influence is not a force, it's a curvature of spacetime, but that curvature of spacetime is caused by the mass of the Sun.

But there's another sense in which, say, Newton's theory and Einstein's theory are more closely related than you might think. Newton's theory contains the problems to which Einstein's theory is a solution. Newton said that gravity travels instantaneously. That was a problem which people recognized before Einstein. They wanted to explain: What does it even mean for something to travel instantly? And then there was the fact that, if the universe lasts forever, as Newton thought, then how come in the long run it doesn't all collapse? And I don't know if Newton was aware of what's called Olbers' paradox: Why is the sky black? But according to Newton's theory, if the universe is either infinite or very big, then the sky should be white. Again, that is a problem Newton's theory can't really answer. You have to make some very ad hoc assumptions to fit that into Newton's theory as a cosmology. And Einstein's theory just solves that problem which was in Newton's theory.

And Newton's theory solves the problem in Kepler's theory, which was so severe that Galileo rejected it. Galileo did not want to believe Kepler's theory because it didn't explain why the orbits were ellipses. If they had been circles, there was [an] explanation that would have fitted into the philosophy of the time. A circle is the perfect shape. If it wasn't a circle, you'd have to explain why isn't it a circle. Kepler was like, "Well, just look, it's an ellipse," and that wasn't good enough for Galileo.

So he had to torture the theory to make it predict circles. But then Newton came along and said, "It's the inverse square law and that can make circles, but it can make ellipses." And that is a deeper level of explanation even than saying circles are perfect shapes. So they're related by their common assumptions, and they're related by the problems that they have or solve.

15:59 **Brett Hall** What you say there, though, it raises the tension between Karl Popper and Thomas Kuhn, who to some extent over-egg this idea that we have these grand revolutions in the history of science that completely overturn the previous paradigm. And anyone working in that existing paradigm is literally incapable of conceiving of how this new paradigm works. Kuhn has a lot more support out there in the intellectual community than Popper, certainly amongst the humanities, even amongst the [sciences] to some extent. And of course, Kuhn has been taken to the extreme ever since by anything calling itself science, like gender science or something that appends the word 'science' to some particular subject. Kuhn did say correct things, but as you just said, it's not the case that we completely do away with the previous paradigm. And the people who create the new paradigm tend to have understood the previous paradigm [in] solving problems from that previous paradigm.

17:03 **David Deutsch** This picture of the young iconoclasts being rejected by the old stick-in-the-muds, and then the young iconoclasts draws together [a] few friends, and when the old-stick-in-the-muds die, then the young iconoclasts become the old stick-in-the-muds. The thing is, it's pure fiction. I don't know of any actual situation where that happened. What does happen is that people often irrationally stick to their own ideas. Whether they are new ideas or old ideas, people can be stubborn. Sometimes stubborn people who support a theory for no reason except that they feel it's right turn out to be right. But there's no algorithm for determining who is right according to who is more stubborn. Sometimes the person who's more stubborn is actually right, like Lister and Semmelweis. They stuck to their guns, they were rejected, but even then it was not a generational thing. There was a much more complex process at work. They didn't just reject a theory, they rejected having to change their working practices that reduced their perceived dignity. But the perceived dignity of doctors is functional, especially in the days when not much was known about medicine. If you told a person that they had to have their tonsils taken out, which was extremely unpleasant, difficult, painful process, you needed a bit of authority, irrational as it is, but the world was much more irrational in those days.

When science got better, people became more open to argument. But the generational story, as I say in [*The*] *Fabric of Reality*, provides no explanation for them changing from one theory to another. It's as if they just invent a new fashion, like when Christian Dior says, "Put up your hemline," then every woman in the world puts up their hemline. It used to happen, apparently. That is not [the] description of what happens in science. There's a reason why people adopt a theory. Even if it's false, there's a reason why they adopt it. If it's not satisfactory to them, they're not adopting it. And sometimes they're irrational. That's just how it is, but it's not a picture of science.

19:19 **Naval Ravikant** I think this is quite obvious if you look at technology. We might have gone from analog attempts at computing to vacuum tubes to transistors, and vacuum tubes to transistors is less of a jump than analog computing to vacuum tubes. Clearly there's progress along the way. Now we don't use vacuum tube computing anymore, it's been obsoleted, but [that] doesn't mean it was wrong. It was a necessary stepping stone. It was closer to the truth, and there was a lot to be learned from there. When you encounter it in real life, then it becomes a lot more tangible and it's harder to refute. I find that the more feedback that you take from other people, the more likely you're to go astray. Whereas the more feedback you take from reality and nature, the closer you are to the truth. And in science, unfortunately, a lot of it gets mixed up in philosophy and academia, where they're not actually interacting as much with the real world. It shouldn't happen in physics, but there is this social feedback loop where you're talking to other people, you're not always building things. The rockets don't have to fly, so to speak.

20:18 **David Deutsch** But the growth of knowledge is possible in philosophy, too, even in morality and epistemology, even when you don't have physical reality. It's this thing I called a few minutes ago—taking the theory seriously. That refutation of solipsism is nothing more than taking solipsism seriously, rather than saying it might all just be my dream. You go on from there, "Okay, if this is my dream, what can we say about my dream? So I'm dreaming the bus, I'm dreaming all the people in it. Now there's a person who is wearing a yellow suit. Did I make that up? I've never thought of it before. Now I'm seeing it." So if I'm a solipsist, I have to have an explanation for how the things in my dream can have come about. And that's really why solipsism destroys itself. And in philosophy, in physics too, most ideas destroy themselves. As you said a little while ago, it's rare to have a case where you can actually decide between two explanations by experiment.

21:23 **Brett Hall** When it comes to progress and understanding, is there going to be a theory that we're not going to be able to understand? I think it's the prevailing view at the moment that there's got to be something out there that is beyond our comprehension.

21:36 **David Deutsch** How do we know that there isn't a limit? How do we know that there'll be no new mathematical knowledge to discover? We can't know. We could be wiped out by an incoming planet from another galaxy that is hurtling through our galaxy at half the speed of light and we'll just be all killed instantly.

There's no known theory that says that isn't going to happen. And similarly, the same could be true in the universe of ideas. There could be a brick wall somewhere where we won't go any further than that. But in both cases, invoking that as an argument about what we can or should do is logically equivalent to believing in the supernatural. Because why did I just say a planet moving at half the speed of light? Why didn't I say an asteroid moving at 99 percent the speed of light? Why didn't I say an illness that operates on principles that we don't know and will wipe us out in a few days? There's an infinity of things I could have said, and all of them make a sophisticated prediction without having an explanation for it.

It's exactly the same when people say that the world is going to end on such-and-such a Tuesday. I would want to ask them, "Why Tuesday? Why not Wednesday?" And they will say, "Because Tuesday comes out of my interpretation of the Bible." And I would say, "Why your interpretation of the Bible and not this other guy who says it's Wednesday?" And pretty much immediately, they don't have an answer to that because they do not have an explanation for their prediction. And it's the same with the idea that the explanatory universality is going to run out for one reason or another, whether [a] physical wipeout or AGI apocalypse or we're all simulations in a computer and so on.

23:31 **Brett Hall** But there is this impulse in people to suggest things like solipsism, the simulation hypothesis, whatever it happens to be, as the final theory. The interesting thing about your work is that you work at the foundations, you go as deep as you possibly can, but at the same time you're against foundationalism. How do you square this circle for people? How do you say, "Well, I'm looking at the foundations, but on the other hand, I'm against foundations"?

23:56 **David Deutsch** It's rather like the relationship between physics and structural engineering. Foundations are theories that explain why the higher-level theories are as they are. But you can't use Newton's theory to build a bridge. To build a bridge, you need theories of bridge building. Christopher [Wren], one of the reasons why he was a successful architect is that he began to use Newton's theory seriously to design buildings. So when deciding what the distance between pillars ought to be, rather than have a master builder's eye for what that should look like and what will or won't collapse, he could actually work it out using Newtonian mechanics. That means that Newtonian mechanics was playing a sort of role of understanding what makes buildings stand up in the first place, and also criticizing particular designs as being not as good as other designs. Then you could use measurement and demonstration and so on to fill in the gaps.

But if you're just given Newton's theory, you wouldn't think of a suspension bridge. Nowhere in Newton's Principia is there a picture of a suspension bridge. That was invented later. So engineering is a separate subject, and you don't study Newton's laws primarily to help you build better bridges. But what Newton's theory did was unify our understanding. It gave us a new level of understanding. It influenced other sciences. People tried to make Newton theories in other fields of knowledge, some of which worked and some of which didn't work.

25:39 **Brett Hall** Now tell me this. Newton, English. Christopher Wren, English. Alan Turing, English. What's special about England? We shouldn't judge one culture as being superior to another. However, it seems as though we've got the beginnings of a special kind of Enlightenment there in Britain, leading to an industrial revolution. What's going on? Why is there so much coming out of England and perhaps the Anglosphere more broadly?

26:03 **David Deutsch** There was the Enlightenment, which largely took place in England, although there were individual people who participated in it in France and Germany as well. But in England, it became the mainstream much faster. It was a rebellion against authority, but it was a nonutopian rebellion. So instead of saying, "Let's get rid of the authority and replace it by the thing that's really true, the thing that is really reliable, the thing that we won't ever have to overturn again." It was a case of, "Look, there's this problem. Some people have privilege, but God tells us that all people are equal. What can we do to fix this problem?" You also had quite rapid social change, economic change, but it all took the form of extending to more and more classes of people privileges that had previously been only in the ruling class. You had Parliament, which was only open to a certain group of people, then it was opened up to more people and so on.

There was a phrase, "The Englishman's home is his castle." Now, I'm not a historian, but presumably an aristocrat's home was his castle, his castle was his home, and his home was his castle, and nobody was legitimately allowed to interfere with him in his own domain. So when you then made reforms that said that an Englishman's home is his castle, that was a modification of existing knowledge of how to structure society. Now you had people who owned houses who were still a small minority, but they weren't the aristocracy. There was a ready-made set of privileges that could be extended until eventually, one after another, they were extended to everyone.

Whereas in France or Germany, it was different. Reforms were all about abolishing things, abolishing the tyrant. To this day, there are traditions of utopianism. The idea is to set up institutions that will last forever, and they are to be set up by fundamental theories like human rights, and you write them down once and for all, then make it difficult to change them and set up institutions that are going to protect those rights forever. But Britain has stuck to its plan over centuries, and it has produced rapid change without any sudden revolutions or without any extremism. In the 1930s, totalitarian theories were very widespread all over Europe, and totalitarian parties either took over or were a major threat to democratic parties. Whereas in Britain, there was a fascist movement, but it never got a single MP, and it went away of its own accord soon afterwards. That's because it was taken for granted in British political culture [that] the political system is here to solve problems. You petition the government for redress of grievances, not to line each other up against the wall and shoot them. The theory was that there is such a thing as a grievance, there is such a thing as redressing it, that it's not easy to do. That the way to do it is to have the rival theories confront each other. You must be allowed to say what you think the problem is and other people say what they think the problem is and so on. Nowhere is it assumed that someone has the final answer.

29:46 **Naval Ravikant** This is why the current rage against misinformation is so troubling. And people even invoke Popper for it. There's a political cartoon that goes around invoking Popper as saying, "We don't tolerate the intolerant, so we have to shut them up because they're spreading misinformation." When nothing could be more the opposite of Popper, which is [that] you have to have debate, have rival opposing theories, have a system for removing bad rulers and reversing bad decisions. In that sense, a clear first-past-the-post system with two parties makes sense because you can hold one accountable against the other. And every eventual successful truth is defined as misinformation by the other side because it contradicts what is already believed to be true. So eliminating misinformation a priori is impossible because knowledge a priori is impossible. It has to be creatively conjectured and discovered.

There is this beautiful idea in *The Fabric of Reality*, and when I try to explain it to friends in my own halting way, it blows their minds. It combines all four strands of *The Fabric of Reality*, talk about epistemology, computation, quantum physics, evolution. If I can summarize the insight, it goes something like this: knowledge is a thing that causes itself to be replicated in the environment. If I figure out how to create fire, then other people in the environment will copy that because it's useful. If there's a gene that is well-adapted to the environment, then the sequence in the gene that leads to higher survivability gets copied, whereas if there's random or junk DNA, that's not going to get copied. And if you look at how the [multiverse differentiates] the randomness, the nonuseful part, the information that is not knowledge, will be different in the [multiverse]. Whereas the knowledge that is useful, the genes that are leading to higher adaptation, the ideas that are leading to higher survivability, the inventions that we're creating that are actually working, the philosophies that we have that are causing us as humans to thrive and replicate, those will be common across the multiverse. So it will almost be like there is a crystal of knowledge.

And I don't think this is doable…If you were somehow able to peek at the multiverse as a single object, then truth would be emergent, or we would be closer to the truth by seeing what is common across the multiverse, and what is different across the multiverse would not be true. This insight, as far as I know, is unique and massively interesting, but is there anything practical out of it someday?

32:10 **David Deutsch** There's a fundamental reason why, even if we could look into the multiverse, it wouldn't be that much help because there is no limit to the size of error we can make. Therefore, when you look around in a multiverse and see all these crystals, yes, on the whole, there are great big fat ones, and you can guess that this one is heading towards the truth. You can't tell where because you don't know where this crystal is going to go. And then there'll be this other great big thing, a religion or something, which has been growing for thousands of years. And there's no way of examining it with a magnifying glass and seeing that it's any different from one that is heading towards the truth.

So we might hope that most of the big ones are heading towards the truth according to some definition of 'most.' In one universe, you can get a hint of that already because you can say, "What idea is most persuasive?" Okay, many bad ideas are persuasive. "What idea is most persuasive to people who adopt it because they think it solves their problem?" Okay, but there are many such ideas that are false, too. So I'm afraid it's not going to work. If there were a limit to the size of error, you would know that, once you've made an error of a certain size, when you have your next idea, it's bound to be true. No one can make more than 256 errors in a row, would be the thing, and nothing like that is true.

33:38 **Naval Ravikant** No shortcuts.

33:39 **David Deutsch** Exactly, there's no shortcut.

33:41 **Naval Ravikant** It seems that the nature of knowledge is that it creates nonlinearities, so even a single false idea can create a false knowledge that overwhelms the truth for quite a while in a large amount of space.

David Deutsch Yes.

Naval Ravikant So it's always creative. It's always conjectural. It's always contextual, which gives an infinity of improvement ahead of us, which keeps life interesting.

ALLISON DUETTMANN:

ON BEAUTY, KNOWLEDGE, AND PROGRESS (WITH BEATRICE ERKERS)

About the interviewer:
Allison Duettmann is the president and CEO of Foresight Institute. She directs the Intelligent Cooperation, Molecular Machines, Biotech & Health Extension, Neurotech, and Space Programs, Fellowships, Prizes, and Tech Trees, and shares this work with the public. She founded Existentialhope.com, co-edited Superintelligence: Coordination & Strategy, co-authored Gaming the Future, and co-initiated The Longevity Prize. She advises companies and projects, such as Cosmica, and The Roots of Progress Fellowship, and is on the Executive Committee of the Biomarker Consortium. She holds an MS in Philosophy & Public Policy from the London School of Economics, focusing on AI Safety.

Beatrice manages the Existential Hope program, encouraging thoughtful discussions on technology's role in society. Beatrice also co-hosts the Existential Hope podcast and oversees its platform, fostering a space for collaborative exploration and ideas. Her work is centred around connecting people and ideas to shape a positive technological future.

Interviewers' X pages:
https://x.com/allisondman,
https://x.com/beatrice_erkers

About the organization: Founded in 1986, Foresight Institute supports the beneficial development of high-impact technology to make great futures more likely. They focus on science and technology that is too early-stage or interdisciplinary for legacy institutions to support. Foresight Institute advances technologies for the long-term future of life, focusing on molecular machine nanotechnology, biotechnology, and computer science.

Organization homepage: https://foresight.org/

EPISODE DETAILS

Date	April 26, 2023
Interviewer	Allison Duettmann & Beatrice Erkers
Source	YouTube
Show	The Existential Hope Podcast
Episode	David Deutsch | On Beauty, Knowledge, and Progress
Link	https://www.youtube.com/watch?v=vYNLahd6fds
Ideas	• Educational institutions are the last institutions of Western society to take on board liberalism and the Enlightenment. Things are taken for granted in schools and universities, which, if translated to society at large, would seem absurd—such valuing obedience and enforcing ritual behaviors. • It's much more important to get right how one thinks that errors should be corrected and the role one thinks that institutions should have than it is to get right the actual policies that those institutions should adopt at any one time. Because if those policies can be corrected, then you can hope that they will be corrected. But if they can't, then you can't. • In science, cranks are valuable. Even scientific publications ought to give some space to cranks. First of all, sometimes they are right. Secondly, even if they are never right, as J.S. Mill said, "You cannot understand the true theory without understanding why the cranks are wrong."
Topics	AI vs. AGI • all problems are parochial • anti-Enlightenment forces today • artificial general intelligence • educational theory • epistemology • existential risks • Hayek • institutions of conjecture and criticism • J.S. Mill • liberalism • obedience • optimism • Popper's political philosophy • problems are soluble but inevitable • Taking Children Seriously • the Enlightenment • universal constructor vs. humans • universal Turing machine • utopianism

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Allison Duettmann** Hi everyone, and welcome to Foresight's Existential Hope podcast. Today we have a very, very special guest that I think is incredibly dear to the Foresight community, and it's no other than David Deutsch. We've really been trying to get you onto this podcast for [so] long because there's a few people that I just really associate with Existential Hope. One of them of course is Anders Sandberg, who we've also had on now previously, but the North Star almost for Existential Hope is you. You wrote a few really fantastic books, including [*The*] *Fabric of Reality*, which is now a little older, but has aged incredibly well. that's on a multiplicity of universes and how that theory, combined with evolution, computation, knowledge, and quantum physics can explain a new worldview.

And then you published [*The*] *Beginning of Infinity*, which really was a big deal for people in this community. You're providing the antithesis to the doomery meme, and you're really saying that, no, progress doesn't have to come to an end. In fact, we're really just at the beginning, and there's a few pretty concrete ways in which we can push progress forward. And also a few more abstract and I think really good memetic pieces [on] how we can think about progress. So this was a really, really great book, and especially Chapter 9 on Optimism has really just stuck with me. If anyone reads anything that I think gets the kernel of Existential Hope across, it's Chapter 9 in [*The*] *Beginning of Infinity*.

And you haven't stopped there. Another talk that is very dear to my heart is, "Why Are Flowers Beautiful?" It's on YouTube, and it's a real treat. And then finally, you're also the creator of one of my favorite child-rearing philosophies. I do not yet have kids, but when I do, they will be raised under Taking Children Seriously, which is the child-rearing philosophy that you and a few other really wonderful minds have put forward. And we also have Chiara Marletto as a Schwarzesemmler Fellow who wrote a really wonderful book on constructor theory, which you both are advancing. And so we're really excited to have you on. So thanks a lot for coming online. I know I said a lot about your contributions as looked at from a Foresight lens already, but if you would like to summarize your perspective on how you got to where you are right now, and your life path a little bit so that people can get a bit of an understanding of what makes you you, that would be absolutely wonderful.

2:45 **David Deutsch** I've never aimed for any kind of global effect that way. Some of the things that I have been interested in have been obviously related. Some of them have turned out to be related to each other and some not. And I don't think one can or should direct one's research, or one's life for that matter, towards a distant, all-encompassing goal. Because that means that if you're wrong, you won't find out until you're dead. All problems are parochial, and if they have universal consequences, that's a bonus. We can be on the lookout for universal consequences, just as we're on the lookout for all interesting consequences. But the main thing is to solve the problems as they come up.

So I'll just give you an example of that. I was interested in quantum computers. I was interested in the theory of computation more generally, and interested in how that relates to thinking. And much later, decades later, came the ideas of an AI apocalypse. Now, it turns out that these other ideas that I had, stemming from a completely different context, make the AI apocalypse look…what's the word I can use? I mean, they are absurd. For a start, if one regards an AGI, something with human-like intelligence but running as a program on a computer, if one realizes that that obeys the same epistemological laws as humans do, then it doesn't make sense to apply different laws of society to it. And especially it doesn't make sense to enslave it, namely causing AGI alignment by force, or building it in into the hardware, as it were. Not that that would be possible, but the attempt to do that is an attempted enslavement. So that's not going to turn out well. And I wouldn't have guessed, when investigating the relevant ideas initially, that it would have any such consequences.

5:40 **Allison Duettmann** I saw you tweet about this a little while ago. Do you find your ideas having any foothold in the AI community? Would you like people to do concrete, specific things differently based on these observations? Is there a particular strand that you want to point people towards?

6:05 **David Deutsch** Well, there are various things involved here. Now, I think that AI, and recently, for example, GPT and ChatGPT, is a wonderful thing and can be very useful. And it has nothing whatever to do with AGI. In fact, as I've written, it's more or less the exact opposite of AGI, because it involves honing the program to conform more and more precisely, and in a shorter and shorter time, to meeting a given criterion. Whereas [with respect to] an AGI, and no one yet knows how to overcome this difficulty, the difficulty is to write a program such that there is no possible idea for which one can say [that] it will never enter that state, it will never have that idea.

Now, people will immediately say, "Well, how do you know it won't get the idea to murder us?" Well, that's the thing. That's the problem that has beset humankind since we have existed. And that's the problem that was solved with liberalism and the Enlightenment. And now we know how to do it. We know how to bring people up in a society that makes it extremely unlikely that they will become enemies of civilization. We haven't got it perfect yet, but we've got it working amazingly well from the perspective of history. From the perspective of history, the fact that we have so few wars, so little violence, as Steven Pinker likes to point out, is unprecedented. And it's not inevitable. It's not that this had to happen. And it's not that it has to continue. It's just that we have the knowledge, both theoretical and institutional, to keep it going as it has been for hundreds of years. And if we continue improving it piecemeal, as Karl Popper would have us do, then there is no known reason why it should stop. But it's not inevitable. It will all depend on what we choose to do.

8:51 **Allison Duettmann** So if you were in the AI alignment communities, would you advocate for [a] Taking Children Seriously view for AI, like [a] Taking AI Seriously, view of actually bringing them up in a specific way?

9:09 **David Deutsch** Yes. In general, the history of educational theory since the Enlightenment has been one of increasing freedom for children and increasing integration of the values of society in general with those of educational practices and institutions. So that has come together. And educational institutions are kind of the last institutions of Western society to take on board liberalism and the Enlightenment. Things are taken for granted in schools and universities, which, if translated to society at large, would seem absurd—valuing obedience, enforcing ritual behaviors, that kind of thing. But this is today better than it has ever been. It is still improving. And I think that if AGI were invented tomorrow, it would indeed be the right thing to do to educate the newly programmed AGI as closely as possible in the way that our society educates children. I think I know of improvements upon that, but it would be wrong to enforce my narrow view of how to do things on everybody. But for everybody to conform to the standards of society at large is not impossible. And to do it for an AGI is not impossible, either.

11:08 **Allison Duettmann** So you would always be arguing for, I guess, more freedom in the way that we educate AGI compared to what the general canon in the AI safety community is.

11:24 **David Deutsch** Well, I don't advocate this for AIs. For AIs, I'm happy for them to be enslaved and to be forced to do whatever we want them to do as accurately as possible. In fact, there is a whole field of making sure they do this, so that self-driving cars don't run over people and that kind of thing. That's all fine. And the more accurately that is done, the better. But that is not how you get people to be members of a free society. You have to do in some sense the opposite. And we have learned slowly and painfully over the centuries to do some very counterintuitive things and to entrench those as fundamental principles of the legal system and of the financial system and everything. We have policing by consent. 500 years ago, nobody could possibly have understood what that phrase means. Government by the people. Nobody would have understood that, either. If you said it to them, they would have imagined some monstrous system which couldn't possibly have worked. But society evolved through conjecture and criticism and cultural evolution to make these things work and for them to become second nature. To throw them away in regard to AGI is terribly dangerous. It is the very danger that the AGI alarmists are afraid of, and they want to do the opposite of what's necessary.

13:29 **Allison Duettmann** Yeah, we wrote a little bit about extending frameworks of voluntary cooperation towards artificial entities. And I think it would be interesting to actually see how those could look like in practice. So, many of the institutions that we currently use to cooperate through in a relatively consensual manner compared to do as you said. It's an interesting theoretical exercise to think about what those would look like in an AI context. But obviously, you don't only have thoughts on AI. You clearly have an incredible breadth of being able to synthesize different fields. And finding really sensible parallels between them. So for a young talented person entering your space, would you be able to give a rough bird's-eye view of what it is that you're working on, thinking about, so that they can maybe get up to speed a little quicker? And I know that in a previous podcast, you actually said that you don't like giving advice. So this doesn't have to be advice. This is just from your individual standpoint. How would you categorize your view?

14:46 **David Deutsch** Yeah. So I also said that giving advice is not a good relationship to have with somebody. "Getting up to speed" is also a little bit misleading, because, although in all the things I'm interested in, there is quite sophisticated knowledge, if you're indifferent to it, you will waste your time or you are likely to waste your time or something like that. But not being indifferent to it doesn't mean getting up to speed. There is no such thing as speed. I think a better metaphor is the one used by my old boss, John Wheeler. He said in physics, but I think it's true of everything, "In physics, every point is a growth point." So wherever you look, even if something has been known for centuries or something has been just invented today, either of those things can be a point of growth where somebody says, "Why should it be like that? What would happen if it wasn't like that?"

And then, of course, most such conjectures are wrong, but they are the means by which progress is made. So I would, if I were starting out now, as indeed, I suppose I am, everybody is, then I would want to think about the interesting things and think about what might be wrong, what seems wrong, what I don't get. Too many people think that if they find something they don't get, it must be because there's something wrong with them. That's not true. If you find something that you don't get, there's almost certainly something wrong with something else. It's either with the people who've told you about it, or the authors of the books, or the teachers of the courses, or whatever, or there is something wrong with the actual material. And even if the material is literally true, they may be looking at it the wrong way. And your perplexity may be, and in some sense must be, the fact that you're looking at it in a way that wasn't intended, and which has some potential for improving it.

17:41 **Allison Duettmann** I guess that is again rather Popperian, which is a nice way to look at your own updating within a field. All right, so as someone perhaps entering your field, they may want to know, have you realized any specific culture shift that [was] relatively instrumental in your life that could either have been throughout your academic career, where the general, canon within your field has shifted, or on a personal level, when were things where you have significantly updated, for example, and were there any specific moments that really got you to update your worldview? Was it relatively stable over time?

18:32 **David Deutsch** Well, I think my worldview has only been largely shaken or shaped once, and that is when I got to understand Popper. But it has been course-corrected several times. And I suppose the best-known one of those is when I decided to update Turing's work on the universal computer, in the universal Turing machine, to include quantum mechanics. And that was after I had realized that Turing had made tacit assumptions in his analysis about physics, and these tacit assumptions were false. And what's more, that these tacit assumptions were now being used in things like complexity theory, to derive what they thought were mathematical theorems, but were in fact consequences of the wrong theory of physics. So they got the wrong answers for it. I mean, I only realized that later, but it turned out that, as a result of making classical assumptions, they got the wrong answers for things like what computational tasks are easy and what are difficult.

20:04 **Allison Duettmann** Yeah. And I think you were actually relatively successful at going out there and at least correcting that error, or at least providing an alternative for that. So that's a great, I think, embodiment of Popper's falsification. He co-founded the, or founded the, the philosophy department at the LSE that I was in. And so it was like Popper up and downNevertheless, I think I only gradually [came] to understand the very critical role that he actually plays in everyday lives over time. You understand someone theoretically, and then over time, it really sinks in as you continue your [life].

20:50 **David Deutsch** That was very much the case for me, too. I mean, when I first got enthusiastic about Popper, my impression of what Popper's theory was, was very wrong. I would now not regard myself at that time as being a Popperian at all, because I'd misunderstood most of the things. Not to put myself down too much, what I had understood was that the conventional way of looking at epistemology and knowledge was just wrong, completely wrong. What I didn't understand is just how accurately and powerfully Popper superseded it.

21:45 **Allison Duettmann** Popper often gets talked about also in context with Hayek as two proponents of the open society. I wonder if you're influenced by him at all?

22:09 **David Deutsch** I think I've only ever read one book by Hayek, *The Road to Serfdom*, and it was all right. I didn't find anything in there that I kind of didn't already think must be true, something like that. Hayek is basically a right winger. So in regard to economics, I agree with him. In regard to society at large, I don't always agree with him. And Popper, I think he overlapped a lot with Hayek, but there were places where they disagreed. And where they disagreed, Popper was usually right, except that he was, to his dying day, I think he was a leftist and Hayek was a rightist. But that only affects their ideas in terms of the color and tenor of their ideas, not so much particular policies, which I think in Popper's case, he wasn't that interested in even.

But Popper's and Hayek's meta take on political philosophy were much closer than the political policies that they actually advocated. And that's much more important. It's much more important to get right how one thinks that errors should be corrected, what role one thinks that institutions should have and that kind of thing, is much more important than the actual policies that those institutions adopt at any one time. Because if they can be corrected, then you can hope that they will be corrected. But if they can't, then you can't.

24:10 **Allison Duettmann** Okay, wonderful. Well, that was just to kind of satisfy my own curiosity. Another question I had is, what, if any, relationship [is there] between Taking Children Seriously and the more scientific work that you've done, what prompted you to go out there and seed this really wonderful movement? Education is just incredibly valuable. And that's how we will shape the future [in] a pretty personal sense. But was there any bark that got you?

25:00 **David Deutsch** I don't think that there is at present, and perhaps there never can be, such a thing as a science of education. I don't think education theory, or even educational psychology, has the potential to be a science even in the future. So it's all philosophy. And for me, Taking Children Seriously is simply the application of Popperian epistemology, and more broadly, liberalism to the foundations of education. It's rather paradoxical, because in a way, that means it's not much of a change. Since liberalism is the kind of dominant assumption in our society altogether. It's completely normal to appeal to things like freedom of speech and individualism and so on in society at large. People may disagree with particular cases, but they won't say, "That's not a way to argue."

But on the other hand, because of meme theory, because of the way that memes work, there is a strong tendency for anti-rational memes to particularly manifest themselves in education. Just like, if you can accept this analogy, it's just like in biology, the parts of our genome that are most resistant to change are the ones that determine the structure and function of ribosomes, and generally of the DNA code. So the DNA code has been almost unchanged for three billion years. It has undergone slight changes, you know, different species have slightly different ribosomes, and animals and bacteria have slightly different genetic code and so on, but it takes hundreds of millions of years for that to change. And that's because the selection pressure on this thing that is involved in replication is stronger than for anything else. So in regard to human ideas or memes, that's the education system or the education practices. Now, this is not the counsel of despair. I mean, memes are not genes, and we are not victims of them, we can always choose to behave differently, and we can always use argument to decide instead of dark feelings that one gets when one does the unconventional thing. So we can, it's just that, it's no accident, I think, that education is the part of society that has been slowest in adopting the values of the Enlightenment.

28:48 **Allison Duettmann** Okay, really, really interesting. Thank you. I also had a question on the chapter on hope that you wrote in [*The*] *Beginning of Infinity*. I think the chapter on optimism is one that really brings the point home in a wonderful way. Because I think one thing that you often get, that certainly, I think, an existential hope lens on the world sometimes gets, [is], isn't this just Pollyanna-ish, and you're entirely ignoring the rift? It seems like you're fighting an uphill battle there by just making a claim that there are good reasons for optimism. So I wonder if you could lay a few out here. Obviously, you can't summarize the entire chapter, and people should definitely go read it if they feel so inclined, but what are a few good reasons for optimism?

29:43 **David Deutsch** So maybe the first thing to say is not good reasons for optimism, but almost like the one thing I have in common with the doomsayers, which is that I don't think anything is inevitable. Human improvement is not inevitable. It is always down to the choices that people make, and there is no limit, there's no naturally imposed, God-given limit on the size of errors that we can make. We can mess it all up if we make the wrong choices, and that conditions how one can become optimistic, or how one can have an optimistic worldview, while being able to combat the objections that you mentioned, that you run into. So optimism is not what I call blind optimism. It's not the theory that things will go right, even though they look as though they will go wrong. Just like blind pessimism is the idea that things will go wrong even if they look good, which also is quite a popular view. It is that, because what will happen depends on our choices, it depends on the knowledge we will choose to create, and on the knowledge that we will not choose to create, and on the ignorance that we will not leave ourselves in. Because of that, there's no reason to give up on any problem. So problems are soluble, problems are inevitable, as I have also said to carve in stone, and also to carve in stone that they are soluble. And they are soluble by specific—not methods, because there are no methods for problem solving—types of process [that] can lead to solving problems, and specific types of process can inhibit the solving of problems.

So conjecture and criticism and institutions of criticism and error correction and of consent are necessary. They are the things that are most precious in maintaining our forward momentum in regard to ideas, because if they are impaired, it impairs everything. And once everything is impaired, well, civilization has collapsed before, and I see no sign of our civilization collapsing. But as I said, there's no supernatural force holding it up and enforcing continued progress. It'll be up to us. And if everyone decides that progress is in fact bad, that progress is in fact an illusion, that progress is always at the expense of one group of people in favor of another, if that becomes a prevailing view, then progress will stop because nobody wants it. And once it stops, there's no reason why it should start up again. Historically, it stopped and it started up again. And in these smaller scale cases that I describe in the book, like Athens and Florence and so on, it didn't start up again. It was just taken on board by the general Enlightenment. But I don't know of any law of nature that says that the Enlightenment had to happen. I think we should be very grateful that it did happen and we should try to keep it going. And we should try to improve it because it still is very flawed. It always will be, always will be very flawed. We will never reach a non-flawed or almost non-flawed state.

34:34 **Allison Duettmann** But is it then that you think that perhaps the biggest risk that we're facing right now is more like distractions [from] those institutions of, conjecture, criticism, and consent that it took us so long to build because we got distracted by some other things that we think are actually higher risk and that the solutions that we try to put forth are actually destroying the [institutions] that took us a long time to build?

35:01 **David Deutsch** I'm not convinced that either that risk or all the risks proposed by the doomsayers are in fact very great. I mean, because they're so important, it's worth taking them seriously, but I don't think the actual risk is very great in either sense. What I can say is that whenever our institutions are impaired by some fad or fantasy or bad idea that's going around, it is bad. People are suffering as a result of every time institutions and traditions of criticism and consent are impaired. People get hurt. People die of it. From the point of view of civilization as a whole, I don't think it's anywhere near that level of harm. But, you know, every child that gets dragged to school against his will is an impairment of the growth of knowledge of civilization. And who knows what has been destroyed thereby.

36:31 **Allison Duettmann** Yeah, that's beautifully said. All right. Well, thanks a ton. I will be handing it over to Beatrice for now. You really have changed the ways that people in this community perceive the world in really wonderful ways. And shows in how people show up to each other and interact with each other in the way that I think oftentimes we are able to hold down critical conversations. And I think if you don't get reminded of these reasons for why that's so important, every once in a while, it's a bit harder to do. So thanks a lot for being so well-spoken and for living in a really wonderful wayAnd I'll hand it over to Beatrice now.

37:33 **David Deutsch** Good to hear. Thank you.

37:37 **Beatrice Erkers** Yeah. Thank you. I'm going to ask you more about the Existential Hope-related questions. There's this sort of idea that's talked about a lot now from Toby Ord and The Precipice, like we're in this very crucial time in history where what we do now has an unprecedented opportunity of shaping what the future in the really long term will look like. Or Holden Karnofsky writing about this being the most important century and we're facing these sort of unprecedented risks. What's your take on this?

38:22 **David Deutsch** Well, I don't think so. First of all, and although nothing follows from this, but perhaps it's worth noting that pessimism throughout the centuries and also conservatism in the bad sense of the word, of opposition to progress, has always included the idea that we are facing an unusual moment of crisis in which the whole of everything we value is at stake. It has always been false, and I think it's false today. I think the talking about existential risks, obviously, you know, there is a risk that weapons we have available today could bring down civilization, though it's a bit far-fetched, but never mind. I mean, they could cause so much suffering that trying to avoid that requires as much effort and attention as avoiding the destruction of civilization altogether or our species. I mean, I don't think I make a distinction there. But we have those weapons and the ancient Romans had enough weapons to do that when they destroyed Carthage. Exterminations and destructions of civilizations have happened since the dawn of civilization. Weapons have been used in unprecedented ways since the invention of weapons.

If anything, I think the amount of knowledge that exists today, and knowledge is not so easy to destroy, that is explicit knowledge. The knowledge in institutions is relatively easy to destroy, unfortunately, but the explicit knowledge is so enormous today that it's hardly conceivable that a civilization brought to its knees could not rise again because they would just have to implement the existing knowledge. They wouldn't have to reinvent agriculture. They wouldn't even have to reinvent the tractor or fertilizer. They would just have to look in a book and it would tell them what to do.

I think that the danger is not as it is painted. It's completely different. On the other hand, the danger from nature is definitely less. So, we've just seen in the last few weeks that a whole range of possible destructions of civilization from a meteor strike [is] not going to happen because technology has advanced to the point just recently where that will not happen. There's still a whole class of possible impact from celestial objects that we do not yet know how to counteract, but a large class of them and the most probable ones, we think, we don't know that for sure, but we think, are now no longer a danger. So, whereas there was a danger of a continental destruction size impact every 250,000 years I think it is, that is now gone. So, one chance of death every 250,000 years multiplied by eight billion people is quite a large risk per person per year. Manifestly existential risks are diminishing.

42:50 **Beatrice Erkers** Well, that's very nice to hear. Also, that's a message I haven't heard in a while. It's our experience that it seems really hard generally for people to envision positive futures, whereas these sort of dystopian futures are easy to see. But you've argued that all problems are solvable and even though problems are inevitable and some are really, really hard, it doesn't mean that they're unsolvable. Have you ever thought about any specific visions of the future that you think are desirable? Do you have a vision of existential hope for the future?

43:36 **David Deutsch** Because of Popper, I think I'm kind of constitutionally opposed to utopianism, both both to utopianism as a philosophy, that is the idea that one should try to design a perfect society and work towards it, and also utopianism in the idea of just imagining what perfection would look like. I would rather look for imperfections in what we have, which, as I said earlier, [are] always parochial, even though they might lead to something universal. But the actual flaw is always parochial, and I'd rather look for those. I have to restrain myself from being the guy who says something's wrong on the Internet, you know, something's wrong on the Internet, so I have to fix it. So I try not to do that. I try to look for things which are going to be interesting to fix rather than just something someone said wrong.

So I think in general terms, I would like the future to be one of ever more rapidly increasing knowledge, ever more rapidly decreasing suffering, but not just suffering in the airy-fairy sense, specific suffering that we see, like people dying of plagues, people dying of pandemics, wars, and so on. These things require a lot of thought, and there's no law of physics that says we can't solve them. Therefore, we can solve them, but it requires creativity. So I envisage the future getting better in ways of involving conquering evils that we know about, but also getting better in ways that we can't possibly know, which will be wonderful.

45:52 **Beatrice Erkers** Yeah, I recall also you've written how creativity is an extremely important tool in gaining this knowledge that you think is what we need more of. We've spoken about Taking Children Seriously. Is there anything else that we should do on a societal level to encourage more creativity and that would enable more knowledge?

46:18 **David Deutsch** Yes. At the moment, Western culture is suffering from a wave of fads whose general theme is to oppose Western culture, Western civilization, to oppose the Enlightenment as I said earlier, to claim that it is fake or that it never happened or that it did happen but was bad, and all that kind of thing, none of which is true. And all of it is based on factual misconceptions as well as philosophical errors. But there is the phenomenon of this informing people's worldviews. There are several such things which are sweeping Western civilization, and all of them have the effect of inhibiting progress by inhibiting freedom, so restricting the range of behaviors that are tolerated for humans, restricting speech and communication so that there are more and more things are becoming taboo.

So all those things are bad. All those things have got reactions against them, which I hope will eventually win or will be replaced by something even better. In this context, I should say that just like I have sometimes said, and people have criticized me for saying, that in science, cranks are valuable. Even scientific publications ought to give some space to cranks, because it's not just that sometimes they are right, like J.S. Mill said, you know, "Sometimes they will be right," but even if they were never right, as J.S. Mill also said, "You cannot understand the true theory without understanding why the cranks are wrong." And not just one crank, but lots of cranks. And I think cranky moral and political theories are in the same category. The danger is, unlike in science, that they get into power and suppress progress towards true theories. That's different. But the cranks, the Woke, or the extremists and so on, are also a source of problems to think about and to apply creativity to. The danger is only that they get into power. That their ideas spread is not in itself dangerous. And our society is good at not letting dangerous people into power. Not infallible, so let's bear that in mind.

50:29 **Beatrice Erkers** Thank you so much. There are two more questions I want to make sure I have time to ask, one of them was on Twitter today. You got a question about how you mentioned that the idea of the universal constructor that you mentioned in *The Beginning of Infinity*, you said that it's flawed. Is that something that you could maybe expand a bit on?

50:58 **David Deutsch** Yes. Well, it's not a very important point. It's mostly a matter of terminology. In *The Beginning of Infinity*, I said that I classified humans as universal constructors, by which I meant that there isn't any fundamental limitation on what we can build or what transformations of physics, physical objects, we can perform if we want to, other than the laws of physics. They are limitations, but nothing else is. That's the point. Now, the thing is, since then, I have actually tried to develop constructor theory in general and in particular the theory of the universal constructor. And it turns out that it is really essential in the theory of constructors, just like in the theory of computers, to imagine objects that obey their program. So a constructor, first and foremost, obeys its program. And then you can ask, what are the range of possible programs that it can be programmed with and what can it do as a result? A universal constructor is one that can be programmed to do anything that is possible to do, to perform whatever transformation doesn't violate the laws of physics. So therefore, a universal constructor must be perfectly obedient.

And a human is almost by definition, like I said at the very beginning of this chat, cannot be obedient. Something which is creative cannot be obedient. So that's a contradiction. Now, you can say that a human body is an approximation to a constructor, because although the mind can't be programmed, it has to consent or at least acquiesce, or then it might fight against what it's told to do and so on, unlike a constructor. But the body more or less obeys the mind. Not perfectly, but well enough to count as an approximate universal constructor. But there's also the fact that humans are very slow at some things. And whether it is possible, we don't know how to make a real universal constructor yet.

But supposing someone designed it tomorrow, it might be something like a computer with a robot. And whether an individual person could build that computer and that robot in a lifetime out of ingredients that were naturally occurring, I don't know. It's doubtful. So there are limitations on humans as universal constructors. But as I said, that's really not very important. It's just a change in terminology from what I used in the book to a more convenient terminology. It doesn't mean that there's any limitation in scope of what humans can do. We don't start with naturally occurring things. If I want to build a physical machine, I will not begin with digging for iron. I will go to the hardware store or to Amazon and buy the things which are close to what I want to make and just assemble them.

55:01 **Beatrice Erkers** Thank you. The second question that I really want to make sure I get to ask you is that, one of the things that we try to do with this podcast is to try to inspire more positive visions of the future. And so we always ask for an example of a eucatastrophe. So basically the opposite of a catastrophe, an event where the expected value of the world is much higher after the event. And so I was just wondering, could you maybe share if you have a vision of what could be such a eucatastrophe? Maybe it's the creation of the universal constructor or something like that.

55:45 **David Deutsch** Yes, I was about to guess that one. I think it will be important. It will mean that after the universal constructor is built, after the first one is built, it can build then more, exponentially more, the human role in production will no longer ever involve toil, that is, unpleasant physical work. Toil will be completely ended by the invention of the universal constructor, although, you know, civilization in general has already reduced toil by something like 99 percent compared with what it was when the human species first evolved. So this is nothing new, but I think it will be fairly dramatic by the standards of everyday events. And the role, instead of being to provide toil, the role of humans will be entirely to provide knowledge either for its own sake or to program the universal constructor. And there will be increasingly sophisticated aids to programming the universal constructor, just like ChatGPT can take a lot of the toil out of writing a program. And all it really does, as I understand it, someone was explaining this to me, is it takes the corpus of all programs that have been uploaded to the Internet and constructs the one you've asked for in the same way that it constructs good English sentences. By the way, I was surprised at how good ChatGPT is at constructing sentences in proper English. I would have guessed that it will be decades before AI can do this. AGI, of course, could do it relatively easily, but I'm not sure that that's on the horizon. I hope it is. But as I have said, people working on this have got the idea that an AGI is kind of, "Just one more heave and our AI will become an AGI." The opposite is the case. The AIs are getting further and further away from an AGI, notwithstanding their excellent English.

58:40 **Beatrice Erkers** Yeah, I saw on your blog you had a bit of an argument almost with ChatGPT about writing a poem. But it got it right in the end, I think.

58:53 **David Deutsch** It did. It does. It often gets it right in the end, precisely when you have inserted in your angry objections all the knowledge that it needs to get it right.

59:07 **Beatrice Erkers** Yeah, well, it was a fun read and I can recommend it. One last question I want to ask. You mentioned Popper a lot throughout this conversation. And if one hasn't read anything by Popper, where should one start?

59:21 **David Deutsch** I'm often asked this and I don't know. It really depends on where you're coming from. Popper was so broad in his subject matter, you know, political philosophy and philosophy of science and philosophy of knowledge, and within those he addressed problems in different ways. I think the concept that maybe unifies all of Popper's thinking in all these subjects, as Matjaz Leonardis recently pointed out to me, is the concept of a problem. A problem in science, a problem in philosophy, a problem in politics. The idea that—and this is also the thing that one of my chats with ChatGPT was about because it didn't know at first, and so I reminded it—according to Popper, the growth of knowledge always begins with a problem. And I asked it, what does the growth of knowledge, according to Popper, always begin with. And it said a theory, a criticism, you know, and I said, "No, it's a problem. Now start again." And finally, it did give quite a nice version of Popper's take on this.

So, to answer your question. If somebody wants to approach Popper, if they've been persuaded by this chat here to start with Popper, to start on Popper, I would say think about what problems you would like to have illuminated by a much, much better theory of knowledge than you have, probably. And that will guide you to which of Popper's books or articles or videos will best make sense to you at first, then later you can see the connections with other things. There's a lecture by Popper called something like "On the Sources of Knowledge and Ignorance".

I'm afraid I can't remember the name, but every so often I go back to read that lecture, it's not very long, and get something new out of it every time. I think it's the best discourse on epistemology ever written. It's incredibly deep and yet incredibly clear. The thing that prompted me to this is that Brett Hall had a series of five videos explaining this lecture by Popper. And he ended up saying, "I'm not sure anyone will want to spend five hours listening to my video." And I said, "It's worth it." But you can also read the original, which is nowhere near that long.

1:03:21 **Beatrice Erkers** Well that's a great recommendation to go out on, and I think we can link the talk in the podcast when we post it. But I just want to echo what Alison has already said that we're great admirers of you at Foresight. And we're very happy that you came on this podcast. I am looking forward to see what our AI generator, image generator will make out of your prompt for the universal constructor. Thank you so much, everyone, for coming. And thank you, David.

1:04:03 **David Deutsch** Thanks for having me.

CHARLES BÉDARD:

THE NATURE OF COMPUTATION, INCOMPLETENESS, AND MATHEMATICS

Discussion host: Charles Bédard

About the host: Charles Alexandre Bédard has a double major in mathematics and physics and a Ph.D. in physics from Université de Montréal, which he obtained under the guidance of Gilles Brassard. During his doctoral work, he explored various problems in quantum computation and algorithmic information theory. Charles is currently a postdoctoral researcher at Università della Svizzera italiana, and he is interested in the interplay between physics and computation; more precisely, he works on the Heisenberg picture of quantum theory, on algorithmic information theory and on algorithmic statistics in physics.

Host Google Scholar page:
https://scholar.google.com/citations?user=L1pflywAAAAJ&hl=fr

YouTube channel:
https://www.youtube.com/@BedardCharles

EPISODE DETAILS

Date	June 28, 2023
Host	Charles Alexandre Bédard
Source	YouTube
Episode	Bennett and Deutsch: The Nature of Computation, Incompleteness, and Mathematics
Description	In this conversation with Charles Bennett and David Deutsch, we delve into the nature of computation, incompleteness and mathematics. The discussion also touches on cosmology, the anthropic principle, the arrow of time, Boltzmann brains, high-level fundamental laws, AGI, probabilities, and quantum theory.
Link	https://www.youtube.com/watch?v=CluVy2jICgs

Ideas

- If fallibilism were not true and there were infallible ways of deriving knowledge, then when we have derived some knowledge, it would never change, and the world would be finite. The world would be a representation of that finite piece of knowledge. Conversely, the real situation is that the world is infinitely amenable to knowledge creation and, therefore, infinitely susceptible to (correctable) errors. But nothing provides a firm foundation, not even the logical rules of inference. Those are all conjectures.
- Often the problem we're trying to solve isn't in data. We haven't yet got any data. Sometimes it's a theoretical problem, like, "How is it possible for Maxwell's equations to be true and geometry to be what we think it is?" There's no data, no data at all. So you have this problem first, then the theory, then the data. And so induction of any kind simply can't exist in that kind of a reality.
- In our culture, it's considered ludicrous to try to infer things about the universe using concepts such as knowledge and computation. But in practice, we already do regard some high-level laws as fundamental, such as the law of the existence of universal computation and the law of increasing entropy.

Topics anthropic principle • Born rule • Church • computability • computation in the universe • conjecture • decision theory • emergent properties • fallibilism • Gödel • Hilbert • incompleteness theorem • infinity • knowledge creation • Landauer • Mathematician's Misconception • microscopic laws and macroscopic laws • probability • proof theory • science • the halting problem • Turing principle • Wheeler

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Charles Bédard** So welcome everyone, I am Charles Bedard, I'll be your host today. I'm a postdoctoral researcher at Universita della Svizzera italiana in Lugano, Switzerland. And today I have the pleasure to be animating a conversation between Charles Bennett and David Deutsch on the nature of computation, incompleteness, and mathematics. I'll please ask everyone who's not David or Charlie to mute themselves so that we don't hear any noise in the background.

0:33 **Charles Bennett** So that means among the [Charleses], that includes yourself, right?

0:37 **Charles Bédard** Yeah, I'll keep myself unmuted. Correct. I'll be animating the conversation for around one hour, and then I'll open it up for the audience to ask questions and jump in. Well, let me introduce our guests to continue. We have Charles Bennett, he's an IBM Fellow at the IBM Research and a Fellow of the American Physical Society. He has been awarded the 2023 Breakthrough Prize in Fundamental Physics for his pioneer work on quantum information.

1:14 **Charles Bennett** Well, it's not just me, what about David Deutsch?

1:18 **Charles Bédard** Yeah, it's a good thing because he's with us, so I'll come up with him at some point. So the Breakthrough Prize was in fact shared between Charlie, David Deutsch, so two of our speakers today, and Gilles Brassard and Peter Shor. So we have two of them with us today. So throughout his career, Charlie played an essential role in investigating and clarifying the roles between, the links between information and physics. Among other things, he coinvented quantum cryptography, he set the basis for quantum information theory, he resolved the Maxwell's demon paradox, and he developed logically reversible computation. Charlie, welcome and thanks for being here.

2:00 **Charles Bennett** All right, good. Glad that's over. Yeah.

2:07 **Charles Bédard** So we also have David Deutsch, a Visiting Professor.

2:08 **Charles Bennett** I think actually Smoluchowski solved [the] Maxwell's demon problem and then people sort of forgot about it for fifty years.

2:17 **Charles Bédard** Okay, good. Yeah. Thanks for the note. David Deutsch is [a] Visiting Professor of physics at Oxford University, a Fellow of the Royal Society and the Institute of Physics. He has also been awarded the 2023 Breakthrough Prize in Fundamental Physics for his pioneer work on quantum computation. David is mostly known for his discovery of the universal quantum computer and the first quantum algorithm, yet he's also a pioneer of constructor theory, and he's made significant contributions to the philosophy of science. David, thank you and welcome for being here. Welcome and thank you for being here.

2:53 **David Deutsch** Thank you.

2:55 **Charles Bédard** So I'd like to start the conversation by discussing the theory of computation. Computers are all around us. Their usefulness in day-to-day life is now evident, but when it comes to computers and the theory of computation, what makes me the most intrigued is what computation is fundamentally and physically, and what role does it play in understanding of the world around us. So Charlie, I'll address the first question to you, but eventually, hopefully, this becomes a conversation. So please, David, if at some point you want to jump in, please feel free to react like you would in a regular living room kind of conversation. Charlie, what is the theory of computation all about?

3:40 **Charles Bennett** You shouldn't ask about computation particularly, but the deep philosophical question is about the relation between mathematics and physics. The way I pose the problem of cosmology is to find a mathematical home for our classical phenomenome. So a phenomenome is like a genome. It's the set of phenomena that we are, the world that we inhabit. Mathematics is a structure which we think is absolute and independent of anything physical. And yet the goal of cosmology or of science in general is to find a mathematical explanation of the world. In other words, to find a part of this vast structure that is mathematics within which what we see is typical instead of surprising and puzzling.

And I think that's why I've gotten interested in cosmology lately, because there are some things that have been learned by modern cosmology that make it harder to feel that we've solved this problem of finding a model of the universe within which what we see is typical. So this touches on the question of the distinction between what is and what could be, and also on the anthropic principle and what questions that some people have asked in terms of self-locating uncertainty. And it also gets into a question that's very ancient in mathematics, which is infinity, the nature of infinity. So some big problems in cosmology. I must say, I always go and ask Andreas Albrecht, because he's been thinking about cosmology for [a] much longer time. And in one of the workshops we organized, he said, "Well, maybe let's take an hour now and wallow in prior probabilities," or "priors." Say, this is a field where you never can solve anything. And one of the main things I got from him was the attitude of being as unsatisfied with your own ideas as you are with everybody else's.

Where was I going with this? So the problem that I said was, what I think one way of trying to approach this is just to think of, as I was saying, infinity is very important. I think we can kind of prove that if you have a finite world in thermal equilibrium, you get a very boring world, essentially because of the Boltzmann brain problem. And so we can get interesting things in an infinite system, which we can't get in the finite ones, qualitatively different. And within that, I would say what I'm trying to find out is how permanent disequilibrium can arise and how, from disequilibrium, complexity in the sense of logical depth can arise and how, in a system that is doing that, science can arise.

And I don't ask about consciousness directly because I think it may be a sort of illusion. So here we have a whole bunch of people who think they're talking to each other on the screens or in the room, but the big philosophical, I guess they call it the hard problem of consciousness of trying to decide what it means that we feel conscious, maybe is asking too much. And it would be easier to say, because our consciousnesses are not independent of one another, if a person is raised without contact with other people, they don't learn to think the way we do.

9:24 **Andreas Albrecht** Charlie, I know it's not my turn yet, but can I just say, is infinity an illusion?

9:31 **Charles Bennett** No, I think infinity is a mathematical notion. But I think that you need infinity to get the kind of permanent disequilibrium that you need to escape the Boltzmann brain problem. You may need it, but you may have another way of doing it. So in other words, I would say that instead of asking questions like, "Is it sort of the Copernican principle? Is it unusual that I'm alive at this moment or that my observer moments are typical?" I would just say, the entirety of terrestrial civilization, or even larger than human civilization, the complexity of the terrestrial world is something that we can look at a little bit more objectively than speaking of consciousness. And so we can say, "Well, maybe whales are conscious in a different way?" And that's not quite a scientific question, but I think we need to look for a mathematical model within which, and I think it has to be infinite—but maybe I'm wrong—within which what we see is typical under a kind of anthropic, as weakly as [possible], anthropic selection. Okay, end of speech. At least, I ran out of breath.

11:06 **Charles Bédard** Is that sort of the reason why you're invoking infinity, because then you could have fluctuations?

11:15 **Charles Bennett** Yeah, if I just have an infinity, let's say like Boltzmann's idea, which gave rise to the Boltzmann brain, of just an infinite universe at equilibrium, then you get the Boltzmann brain problem. You get the fact that you can't believe anything that you see, because everything happens somewhere. And I think this problem is somewhat better understood in the more modern cosmologies. But I was looking at very simple cellular automata models in which, by making the model infinite, and its dynamics is reversible, you get something that looks like unbounded complexity that goes on forever, but it only goes on forever because the model is infinite, and therefore it can never equilibrate and get boring.

I think I've got a picture of it. I'll see if I can find an example of the model like that here. Okay, so this is a reversible cellular automaton, and the time goes horizontally like that. And there's three domains here. It's one-dimensional automaton, so this is [a] time history of it. And these differ just by the reversal of black and white. So there's domain boundaries that collide here, producing this complicated thing that just goes on forever, getting more and more complicated. So this is the sort of thing where… you can't get this in a finite model. If you do it in a finite system with local interactions, it gets interesting. And then if you run long enough, it just gets boring. So that's why I think infinity is important, but that's just maybe I haven't been familiar enough with cosmology. Let's see what Andy says, but I think that the cosmologies tend to be infinite now also. Now I have to stop sharing my screen.

14:19 **Charles Bédard** I was expecting Charlie to throw us into cosmology, but to be honest, I didn't expect on question one [that] we would go onto cosmology. Thank you, Charlie. I enjoyed it. David, would you like to react? Maybe you can either react to Charlie's comments or perhaps you can backtrack the question and retake the theory of computation.

14:42 **David Deutsch** Well, I think that I can react and respond to your question at the same time, because I think where I disagree with Charlie would be right at the beginning, sort of foundationally.

14:58 **Charles Bennett** Oh yeah, you remember, you should say the three R's: react, respond, and refute.

15:04 **David Deutsch** Right, yeah. Simultaneously to save time, yes. So if you ask, "What is computation? What is computation in the universe? What are computers? What is that all about?" If I was answering that question, I would have to have 'universal' and 'universality' right in the first sentence somewhere, because

I think that's what it's all about. Computation is physical, I'm sure we can all agree on that, but I think there is still a hint of the Mathematician's Misconception in Charlie's conception of computation, computers, physics. So I would say computation exists, universality exists in a particular property of the physical universe, namely that the physical universe can accommodate machines which are universal in the sense that they can be programmed to mimic any other machine if they run long enough and have enough memory. And for them to run long enough and have enough memory, they have to be maintained. They have to have additional memory added, and so there has to be, implicitly, there is a background of knowledge creation. If there is to be a sufficiently powerful computer like [a] Turing machine or something, then in the background, there has to be knowledge creation. Just like, I think I first read this in Charlie's thing, like what would happen if you found a fountain pen on the Moon, then it would tell you a massive amount more.

17:05 **Charles Bennett** It would tell you that it'd been settled by people from the nineteenth century.

17:10 **David Deutsch** Yes, yes, most likely. And it's very, very unlikely that it formed spontaneously. And so the most likely explanation for it would be that there had been a civilization, you could infer a lot about the civilization by, as you say, it's a nineteenth century-type civilization, and you could analyze the ink and therefore see what kind of squids they had there, and so on. Another thing that this illustrates, which you didn't mention, is that there's an intimate connection between simple emergent properties and simple microscopic quantities. In fact, there are laws about emergent properties, and the laws about evolution and knowledge and so on are among the ones we would use to analyze the origin of this fountain pen. So we would gain a lot of knowledge about microscopic information from some macroscopic information and macroscopic laws.

I haven't mentioned mathematics. And the reason I haven't mentioned mathematics, even though I totally agree that there's a mathematical world that's sort of super-infinitely large, and that world contains all possible functions and all possible laws and all possible mathematical objects, the vast majority of which we cannot even describe. The ones we can describe are basically the ones that are computable by these machines, which are universal within our universe. So it's no good trying to explain that via what is going to be likely, what you are going to be likely to see, because what is typical, what is probable, and so on, those are all determined by the laws of physics. There is no mathematical notion of probability that applies to physics, unless we have a law of physics saying so. And in some cases we do, and in some cases we don't. And where we do have a probabilistic theory of something, the anthropic principle alone doesn't contain any information about what the laws of physics are.

This is an argument that I got from Dennis Sciama a long time ago, with his reaction to Brandon Carter's famous paper about the anthropic principle. The thing is, if you think of this as, "How do we fix what the dimensionless constants are?" or just the laws themselves, you could think of the laws themselves as being enumerated like in Solomonoff induction, so we have all possible laws and so on. And you ask, "Given what we see, what is the most likely?" That is all inapplicable to fundamental physics, at least not via the anthropic principle, because if you think of the set of all—or the class, or whatever it is—the set of all universes consistent with something, like something consistent with the fountain pen on the Moon, or consistent with what we see, or whatever you want to say, consistent with the existence of computers, almost all of them are very near the boundary of that set. The larger the dimension of the set, the more of it is contained near its boundary. So if the only reason why we're here is that we're anthropically selected, then it's overwhelmingly likely that we're going to die in the next nanosecond, or picosecond, I suppose. It depends how fast the chaos is going to come in on us. And since we haven't, that theory is hereby refuted. So that's not the way to get the answer. I think the way to get an answer is not to try to derive microscopic properties that will then give you the desired emergent properties. It is to think of emergent laws, which, among other things, will give you microscopic laws as well as microscopic initial conditions like the ones that produced that fountain pen. You see, Charlie, your example has lived with me for decades and has changed me deeply.

22:30 **Charles Bennett** Well, I was thinking of an ordinary ballpoint pen, but you've put it back a century.

22:36 **David Deutsch** It's evolved into a fountain pen in my mind because of its deep significance.

22:44 **Andreas Albrecht** The ballpoint pen would just mean you have a hole in your pocket.

22:49 **David Deutsch** "It fell out of an astronaut's pocket" is a likely explanation.

22:55 **Charles Bédard** For you, David, an example of an emergent law that constrains also the microscopic laws would be, "there exist universal computers"?

23:03 **David Deutsch** Yes.

23:05 **Charles Bedard** Okay, would you like maybe to expand a bit more because you've contributed quite significantly to the Church-Turing thesis. I think there was quite a bit of ambiguity of how we should understand the Church-Turing thesis. And now you name the physicality of the universal computer. I right away see a link. Maybe it'd be good to expand a little bit on that, especially for the audience here.

23:25 **David Deutsch** Okay, this is controversial, and I don't think I did. I think it was Turing. So the way the story is usually told is that Turing, Gödel, Church, and Post all kind of came up with the same idea at the same time, all proved the same theorems basically, although they had different points of view. None of them copied from each other, although some of them knew about each other. They kind of converged on a conception. But I think Turing's conception was different. The other three all had the Mathematician's Misconception. They all thought that they were doing a piece of mathematics or, as Roger Penrose would say, metamathematics, theory of proofs. But Turing, the way he solved the Hilbert's decision problem was to think of a physical model of proof, and then argue implicitly, explicitly—that again is controversial—that that argument is decisive. That is, he thought he was proving, and he almost did prove, that no physical object can escape being part of the universal Turing machine's repertoire.

And the universal Turing machine is an idealized physical machine. It's really got nothing to do with mathematics. And its physical relationship with the rest of the universe is what he elucidated. And then he said that that solves the decision problem, because we must regard the decision problem itself as a question about the physical world, including mathematicians and so on. And so that's why he used language like, "functions that would naturally be regarded as computable." 'Naturally' I always took that to mean 'straightforwardly computable in nature'. But when I said that to mathematicians, they howled in rage and said that I'd misunderstood it. Well, if I misunderstood it, that misunderstanding turns out to be the truth. And Turing's writings look exactly like that truth. So I think he knew. He just wasn't used to talking about physics. He was a mathematician, but I think he did not have the Mathematician's Misconception. We're all brought up to have it. We have to free ourselves.

26:38 **Charles Bennett** What is the Mathematician's Misconception?

26:43 **David Deutsch** Well, for present company, I can put it like this. It's the idea that the integers and the logical operations like 'and,' 'or,' and 'not' are given to us by God. They are the natural things. And if you can make some, build some structure or form some conclusion on top of those, then you have proved it. Whereas the real truth is that the integers and the logical connectives and classical rules of inference and all that, they are all given to us by physics alone. There is no underlying mathematical substrate that we can appeal to. If the laws of physics were different, then we could have laws of physics that didn't mention 'and,' 'or,' and 'not,' or integers or real numbers. They could use any mathematical objects anywhere in the mathematical world. But the fact that they don't is a feature of the laws of physics.

With this misconception comes some other misconceptions such as, for example, that simplicity is somehow defined independently of the laws of physics. And infinity as well. As I have written, Zeno was puzzled by the fact that there's an infinite number of points between here and the other side of the room. And how come he can go from one point to another? How come he can do an infinite number of things in a finite time? Well, the answer is that what is finite or infinite physically, its relationship to mathematical finiteness or infinite is a matter for the laws of physics to determine. And they happen to say that this particular infinite thing in classical physics, the continuum, can be traversed. An infinite number of steps of traversing it can be finitely performed. So what can be finitely performed or not, or conversely, what can only be infinitely performed, i.e. can't be performed, is mandated by the laws of physics and not vice versa. So it's physics that tells us the difference between finite and infinite, complex and simple. And also probability, which is a sort of scam lodged in the middle of all these misconceptions.

29:47 **Charles Bennett** Is it a meta-misconception?

30:01 **David Deutsch** There is simply a misconception that probability is another one of these concepts that we can help ourselves to, and we haven't yet invoked any physics. We would like what we see to have at least 0.83 probability, then we won't worry that it's strange. And that invocation of probability we take to be sort of harmless. We don't see that there's a rich world of physics defining what we mean by that. And in the case of probability, unlike the other things, it doesn't even apply to most of the universe or the multiverse. It just applies to very special situations. So there's no excuse for it.

30:53 **Charles Bédard** Let me ask a clarification, David. If I would have been asked, "What's the [Mathematician's] Misconception?" In a nutshell, I would have said, "It's the idea that proof theory is a branch of mathematics, but proof theory is a branch of physics, ultimately." It's given rise to computer science. Computer science has its roots into what are computers. Computers are physical objects. Now, the story you just gave us, you started speaking of natural numbers and logical connectors. How can I bridge those two pictures of the [Mathematician's] Misconception?

31:29 **David Deutsch** Well, they are the same, I think.

31:36 **Charles Bédard** Maybe I'm giving a try and then tell me. Because we typically do proofs with typical logical connectors and that will give rise to somehow our computers, so that we end up abstracting our computers with those logical connectors and those integers. And that's where you say that taking this as a way to formalize proof is basically physical because I'm already abstracting my physics of the computation.

32:06 **David Deutsch** Yes. I think that's basically what I said. I'm not really used to speaking in mathematical language. So I speak of it in terms of laws of physics. So I think, probably, Turing never said this, but I'm sure if he was here now, he would agree that proof theory is a branch of physics.

32:39 **Charles Bennett** Okay. I found myself agreeing with almost everything you said until you started talking about the Mathematician's Misconception, which I don't think Turing would have agreed with you about, and which I think is actually a dysphemism. It's an idea that you don't like, but you haven't shown that there's anything false about it. In fact, I think Turing thought more about physics than his contemporaries, but I think he felt that the discovery of universality meant that the physics that we have here could be simulated by a computer. And the physics that might exist in a very different world, perhaps one with a different number of dimensions or one [that does] not even [have] a notion of locality could also be simulated by [a] Turing machine. In other words, I believe he was very aware of physics, but he didn't think that different parts of the multiverse, think of string theory, the vast number of different supposed things that can come out of string theory, of which we might be one, that they would have a different mathematics. I think he believed in the universality of mathematics and its ability to simulate physics. So that's where I think that you're even calling it a misconception is a misconception of what Turing thought and also a dysphemism rather than any kind of refutation.

However, almost everything you said at the beginning, I agreed with. I said I take for granted that [the] universality of computation, the unsolvability of the halting problem, which is the same thing, is the heart of what the notion of computation is. And that the other thing that I also didn't even say, because I believe it so strongly, is that this mathematics, this universal mathematics of computation is capable of simulating physics, not only the physics that we have here, but if you talk to Andreas, if he's still around here, there he is, yes. In many models of cosmology now—imagine that there are inaccessible parts of the universe, which are no less real than ours, except that we'll never hear about them, in which the laws of physics, such as the number of spatial dimensions and the number of temporal dimensions, if you want, and are extremely different. And in almost all of them, universal computers don't exist. They exist physically, but they don't exist, there's nobody there to complain about it or to celebrate it, but they're there. They're like the desert areas of the universe. But all of that could be simulated by mathematics. So I don't think there is a Mathematician's Misconception. There is just the idea which has been around since Galileo, that whatever physics is, is discoverable by experiment and modelable by mathematics.

36:37 **David Deutsch** I think that's clearly untrue. If by 'universe' you mean the set of all, or potentially the set of all possible mathematical entities instantiated as physical objects, as you say, most of those would not be simulatable by a Turing machine.

36:57 **Charles Bennett** No, no, no, they all would be simulatable by a Turing machine. Most of them wouldn't give rise to the physical possibility of [a] Turing machine in that part of the universe. For example, I don't think there are a lot of Turing machines in the middle of the sun.

37:12 **David Deutsch** But in the space of all mathematical objects, there are objects which solve the halting problem. And that such an object cannot be simulated by a Turing machine.

37:30 **Charles Bennett** Oh, yeah. Yeah, I have. In other words, like the Kleene hierarchy of all of these higher [levels] of unsolvable problems. The problems that you could solve if you had an oracle for the halting problem. Now, I have an idea that I'd like to run by mathematicians, such as yourself, even these mathematicians, which, just as I have been lately over the last five years, seduced by cosmology and so think, "Oh, it must have the answers to all the things I worried about since I was five. Let me try and understand it better." So, you know, you're a mathematician and you've gotten to be so fascinated with physics: "Oh, the answers to all the things I'm wondering about in mathematics must be in physics somewhere. Let me learn some physics."

Well, anyway, this is an idea that you probably know more about as a mathematician. But when I was writing about the Chaitin's omega number, the halting probability of a universal computer with self-delimiting programs, where if you had this number, you would have an oracle for the halting problem, but it would be a painfully slow oracle. In other words, in order to answer any question, you would have to run for a busy beaver amount of time. And I said, "Well, this oracle would be universal in the sense that it would decide all finitely refutable propositions." In other words, problems such that you can express them with one quantifier over the natural numbers. In other words, there exists a time such that if you run this Turing machine for this amount of time, it halts. And then I speculated that the harder problems, which would involve two quantifiers or more, like the twin prime conjecture, might be by and large not interesting, not mathematically interesting, because they could be decided by a stronger but finitely refutable proposition.

In other words, instead of saying there are infinitely many twin primes, you could say that the spacing between twin primes grows more rapidly or more slowly than a certain function. So, in other words, something that has more than one quantifier might turn out to be a consequence of something that has only one quantifier. So what you're feeling about when you say 'natural' in mathematics, maybe these things that are harder than the halting problem are just kind of boring, because probably most of them would be…what do I mean by 'probably'? I'm invoking your scam of probability. Probably most of them would be decided as cases of the halting problem.

40:58 **David Deutsch** I was sure for most of what you've just said that you were addressing Charles, because I'm certainly not a mathematician. Them's fighting words, you know.

41:07 **Charles Bennett** Okay, okay.

41:08 **Charles Bédard** Well, if you have a tentative answer, David, go for it. But I also thought about it.

41:16 **David Deutsch** Go ahead. It's better than mine.

41:19 **Charles Bédard** I remember that piece, Charlie, you wrote in which you say that you specifically spoke about the twin prime conjecture, but would that naturally carry over to higher elements of the arithmetic hierarchy?

41:37 **Charles Bennett** I think so, because if you bound these quantifiers, you can make them go away. And if a bound exists, it could be that there's a proposition that decides the twin prime conjecture by proving a stronger conjecture that is just of the same form as a halting problem. And if you have three or four quantifiers, I believe the same thing could happen.

42:03 **Charles Bédard** Okay, so another sort of possible glitch that I see is, if the spacing in the twin prime conjecture scales, suppose, larger than busy beaver, then no algorithm can basically...So it could be still be true, but unrefutable.

42:18 **Charles Bennett** Certain things of that sort could exist. So I guess this is like experimental mathematics, but you can't do the experiments. I don't know, has anybody thought about that? How do you decide how plausible it is that one of these higher-level conjectures could be decided by a provable...Oh, yeah, it's really a futile gesture, because it's like, if I had a solution to the halting problem, then maybe I could solve the twin prime conjecture without doing any extra work. But that's like, "if my aunt had wheels, she'd be a trolley car," because I don't have a solution to the halting problem.

43:11 **Charles Bédard** Yeah, but it might also be that you had a solution to the halting problem, and you still can't.

43:15 **Charles Bennett** And you still couldn't, yes. So how do you compare the likelihood of those two things? And I'm sure somebody's thought about that.

43:21 **Charles Bédard** And then what we mean by likelihood.

43:22 **Charles Bennett** Yeah, shut up.

43:25 **Charles Bédard** Well, actually, we could move on on the topic of incompleteness, and I think you bring it on. I had a bit of a context maybe to get people into it, but the incompleteness of mathematics was…perhaps we have to start it back to Hilbert, where he was hoping to put into a single formal axiomatic theory, all of mathematical truths. And Gödel, in 1931, basically put an end to this hope by finding a statement that is true and has no proof. But Gödel's statement is self-referential and it might look like exotic kind of statements. So it's tempting to think that these kinds of statements are an anomaly and that we can safely ignore them and keep doing proofs and mathematics like we usually would have before Gödel's result. And I assume most mathematicians live their life this way and I don't blame them.

But seen from an algorithmic information theory perspective, notably due to the work of Gregory Chaitin, one of your former colleagues, I believe, Charlie, incompleteness is a much more widespread and inevitable phenomenon. So, for one thing, Chaitin's result of incompleteness does not involve self-referential statements. And also, they're usually cast in a way that he comes up with an infinite family of true but unprovable statements. And then you say, "Oh, I will enlarge my formal axiomatic theory to be able to prove more of those statements." But by doing so, you only [manage] to prove finitely more, but you're still in front—regardless how big your formal axiomatic theory grows, you're always in front of an infinity of true yet unprovable statements. These are generally the ideas from the algorithmic information theoretic proofs of incompleteness. So, yeah, I think algorithmic information theory makes the case for…incompleteness becomes somewhat natural and actually inevitable.

Charlie, would you like to react or comment or expand on the limits of our formal axiomatic theories? Notably, I have in mind, like maybe some of how we should behave with respect to formal axiomatic theories in the light of these incompleteness theorems. Chaitin has suggested maybe we should acquire new axioms based on their fertility of the consequences that the problems that they help us [solve]. To me, there's a striking similarity now between what theoretical physicists do, where they change or update their principles based on the problems that we're here to solve. So do you have any remarks to make on that line of philosophy of incompleteness?

46:14 **Charles Bennett** Well, I'm more of a physicist. I really love computational universality. This reminds me of what Danny Greenberger said about quantum mechanics. He said, "Any God that would know which slit the particle went through, I wouldn't believe in that God," of the two-slit experiment. So I would say: "Anyone who would like to live in a world that Hilbert wanted to find, I wouldn't wanna be friends with that person." In other words, I think that the universality, and the dual to it is the incompleteness, is just a beautiful feature of the way [the] world is. It's beautiful in mathematics in the same way that quantum mechanics is beautiful in physics…or general relativity.

47:36 **Charles Bédard** Nice, thanks. David, what is your take on incompleteness? And it doesn't have to be from the algorithmic information theoretic perspective, just what are your thoughts?

47:45 **David Deutsch** I would agree with what Charlie just said. The world that Hilbert envisaged is a world without creativity. And probably ultimately, if you take our world and hobble it down to that level, life could never have evolved. So ours is a world in which life can evolve because there's incompleteness in the mathematics that describe our world. I said mathematics just now as a concession to you all. But what I meant is entirely a physical thing. It's a physical property of the world, which is responsible for the possibility of life, and then later, presumably of intelligence and creativity. And science and of the limitlessness of science. Let's put it the other way around. If science was limited, if there was a feature of the universe that limits science in the way that Hilbert wanted to limit mathematics, then there couldn't have been any science in the first place. And this is another one of those connections between what we call emergent properties and what we call microscopic properties.

I would like to see a way of formulating the laws of physics that doesn't discriminate between microscopic and emergent laws of physics. A lot of the sort of notorious problems we have, like how to define entropy and the arrow of time and so on, I think are just because we insist that the world must be fundamentally made of microscopic laws. And so just to let you all know, constructor theory is an attempt to have a scale-independent way of describing laws. I think someone's raised a hand, is that?

50:21 **Sam Kuypers** Okay. So you said that if there was anything that limited science in the universe, if the universe somehow limited the scope of science, then science couldn't have gotten started in the first place. But I can imagine that the observable universe is actually all there is—that beyond the observable universe, there's no matter, there's only space. So there's nothing to help us form larger and larger computations. So at some point we reach the limit of what we can do with computers, or something like that. And that would still be consistent with us [having the ability] to do science right now. I was wondering if you think it's wrong that maybe even if beyond the observable universe, there's no matter, that you can somehow still have science that progresses infinitely, or if [this] is another mistake? I'm curious about the response to that.

51:23 **David Deutsch** Yeah, so I think that is wrong. And it's for exactly the same reason that I was just saying. If we characterize all possible laws as being like our world, but it only lasts a million years, or like our world, but it only lasts a billion years, then there are lots of possibilities where it would look like our world was the actual one. And similarly for mathematics, it might be possible for proofs to reach up to a million steps but no more, because it so happened that the laws of physics make all computers decay after they have performed a million steps. And there are many more of those worlds than there are what we think the actual world is, with no limits. I won't say it's because the limitless one is simpler, because that would be falling into the Mathematician's Misconception. But I want to say something like that. I want to say that the limitless one is a good explanation in high-level terms. If you try to translate that into low-level terms, you'll get to variations of it which look like, "Well, it only lasts a million years, it only lasts 2 million years," and so on.

But at the high level, variations of it are very difficult to find, because you would have to say, okay, proof is limited by…by what? You would have to have some high-level thing that can fit into the language of talking about proofs and limits and completeness and so on. So from the point of view of physics, these are all high-level macroscopic constructs. But I think those are the fundamental concepts in which the laws of physics actually are expressed. And because of that, we can say that insisting on describing it all microscopically is perverse because it's much less simple. But what we really mean is that there are no good explanations along that route.

54:27 **Sam Kuypers** The way I understand what you just said is that it could be that the universe ends at what is observable, that the observable universe is all there is, but that would ruin other explanations that we have, like explanations of how science works.

54:48 **David Deutsch** Yes, if you're going to settle for that, you might as well settle for the fairies at the bottom of the garden. We needn't have embarked on this great project of mathematics and physics.

54:59 **Sam Kuypers** Yes.

55:03 **Charles Bédard** Thanks. David, would you relate the incompleteness phenomenon in mathematics with fallibilism in knowledge creation in general? It seems like you sort of invoked it when you invoked life. Is that [a] similar idea?

55:19 **David Deutsch** Yes, it's again a similar idea and for the same reason. If fallibilism were not true and there were infallible ways of deriving knowledge, then when we have derived some knowledge, it would never change and it would be true and the world would be finite. The world would be a representation of that finite piece of knowledge. Conversely, the real situation is that the world is infinitely amenable to knowledge creation and, therefore, it must be infinitely susceptible to errors, which, however, can be corrected. But nothing provides a firm foundation, and I repeat what I said earlier—not even logic, not even the logical rules of inference. Those are all conjectures. We pick them because they seem right, because they seem useful, they seem fruitful, but who knows, we may find new ones.

56:38 **Charles Bennett** Are you saying that there could be a world in which pi was rational or square root of two?

56:48 **David Deutsch** Well, of course, that's a fairly simple case. All it would have to do is have a different geometry from ours, not Euclidean geometry. Pi isn't instantiated in the universe, anyway. It's an idealization. So if you actually measure a circle, you'll never get pi as the ratio between the diameter and the circumference.

57:20 **Charles Bennett** I was thinking of the mathematical pi.

57:24 **David Deutsch** Well, so there are two different things one might mean by the mathematical pi. Physics allows us a certain window onto the class of abstractions, a tiny window. We can see some of them. We can form theories about some of them. We can learn about some of them, and they include Euclidean geometry with its pi. That pi cannot be changed by anyone, not even God. Not even if we look out of a different window. But if we did look out of a different window, thanks to having different laws of physics, or if the laws of physics were not what we think they are but are a bit different from what we think they are, as happened with Euclidean geometry and general relativity, then we won't find pi out there in the stuff we can look at. We can still describe it. We can describe pi, and we can describe the universe as it actually is, which doesn't have pi. And there will be other ones which we can't describe, but they may have people in them which can describe them. And then there'll be infinitely more where there aren't people who can describe anything. And be careful to conclude that therefore we don't exist, because that is like the anthropic principle misconception.

58:53 **Charles Bennett** You love misconceptions. I would say misapplication of the anthropic principle. These places exist, but there's nobody there to complain about it.

59:03 **David Deutsch** No, but you want to use it to deduce something about the world we do see. And I don't think the anthropic principle is powerful enough to do that.

59:20 **Charles Bennett** Yeah, I agree. I want to say something about on the edge, because you reminded me of that, and I want to show you a picture. Is that okay now? Okay, so I started worrying about this in terms of the Boltzmann brain problem, which most of you are familiar with, but here's a sort of a summary of it. The New York Times version from Sean Carroll. This was Boltzmann's idea—that the reason the universe is out of equilibrium is anthropic. That is, if the universe is infinite and at equilibrium, but we couldn't exist in one of these typical parts. So we're in an atypical part, anthropically selected. But then somebody else, I think it was Eddington, said, "Wait a minute, if that's true, then we most probably are in the smallest fluctuation consistent with our brain existing." And so the idea was that you would get something like this.

So people worried about equilibration, and where in the nineteenth century, they called it the heat death of the universe, but they thought it was a different problem for the distant future. But Boltzmann or Eddington showed us a problem in the present, undermining our ability to make inferences about conditions in the past or elsewhere, because the inhabitants of any universe that will eventually equilibrate would have to make the additional postulate that they're situated atypically early in its history. Now that gets into David's very nice pointing out that from [an] Occam's razor point of view, a world that is just like our world but ends in a second from now or a million years from now, is almost as simple a description. So anyway, the Boltzmann brain problem says that if the world equilibrates, then almost all places that experience the same phenomena that we have, these are illusory phenomena that don't give grounds for scientific inference.

So now this is a little bit like the doomsday problem, which you can argue that the world that has civilization in it has only existed for a very short time compared to the time available to it, and why are we so atypically early? Maybe it's because civilization is intrinsically unstable, it'll destroy itself, or maybe there's something that I think David would like—perpetual newness. That is, maybe a billion years from now there will still be people, but they will be preoccupied that some qualitatively new feature of their existence, which they consider very important, is so new and they wonder why they are so near the beginning of infinity, as David would put it.

So I would like to credit the anthropic principle to Schopenhauer, who really expressed it in the nineteenth century, 1844, before Darwin. And here he says that the world is on the brink of self-destruction, and we should expect to find it that way because of this surface-to-volume argument in high-dimensional space that David made. That is, that if there are many variables and all of them have to be within a certain range for the world to be habitable or for it to support life or for it to support universal computation, then with the highest probability we're right near the edge and only a little, a very close to self-destruction. And this is Schopenhauer's, he didn't put it in mathematical terms, but if you did, if he says we do not live in the best of all possible worlds, in fact we live in very nearly the worst of all possible worlds, because if we imagine goodness of a world depending on many parameters here, too, and we just take everything to lowest order, and the goodness is a quadratic function and the best of all possible worlds is the maximum, and then we find that we're very close to the edge, so we're probably right around the edge here. Therefore, we should expect the world to be [on] the brink of self-destruction, as it apparently is politically right now. And so we just have to hope that it survives. Well, that's my comment that I thought had to do with what we were talking about.

1:04:45 **David Deutsch** Yeah, again, I advocate thinking in terms of high-level fundamental laws rather than insisting...

1:04:53 **Charles Bennett** How do you discover these high-level fundamental laws, how do you discover them or refute them?

1:04:59 **David Deutsch** The same way that we discovered low-level ones: conjecture. We have a problem that we think might be soluble by postulating a law of nature, and the law of nature seems to answer that problem and many other problems. And so if it doesn't, then we haven't solved the problem yet. And the more it does, the more problems it exposes, which we then solve and so on. Now, we're doing this at many levels already. There are people who have deep theories about what it takes to win a war and what it takes to make a stable society and what it takes to cure depression and so on. It's just that we have a culture that stigmatizes those from the point of view of being fundamental. So we only expect such knowledge to extend to our own planet, our own time, our own species, if that.

We don't think of them as being fundamental, but some of them, like the law of the existence of universal computation, and the law of increase of entropy for that matter, they are, from a practical point of view, we do regard them as fundamental. Like as Eddington or someone said, "Somebody tells you that the first law of thermodynamics is wrong, then so much the worse for the theory, but if they say the second law is wrong, then they must retire in deep humiliation," and so on. So we do actually have confidence in high-level laws. It's just that we have a culture that tells us that those can't be fundamental. We can't kind of infer things about the universe and the Big Bang and the long-term survival of the universe. And bits of the universe that we can't see. It's considered ludicrous to try to talk about that using concepts like knowledge and even computation and information and so on. Information actually is a bit of an exception because, thanks to the Mathematician's Misconception, people are inclined to think that information is fundamental and everything might be made of bits.

1:07:54 **Unknown** So you like "bit from it" rather than "it from bit"?

1:08:05 **David Deutsch** Yeah, absolutely. Yes.

1:08:07 **Charles Bennett** This is very much Landauer's [view]. You probably sympathize with what he said.

1:08:13 **David Deutsch** About this?

1:08:15 **Charles Bennett** Information is physical.

1:08:17 **David Deutsch** Oh yeah, of course, I thought we all did. I [thought] that's his great contribution to the world.

1:08:24 **Charles Bennett** Wheeler put it the other way. He says physics is informational.

1:08:28 **David Deutsch** Yeah, well, that's a misconception. That obviously can't be right. It's the same thing as expecting us to be in a simulation and the aliens simulating us on a giant computer. And, for some reason, people think that that computer has got to be a Turing machine. That's simply a non sequitur. It's just a parochial forcing of a human concept onto imaginary superhuman aliens.

1:09:06 **Charles Bennett** You mean you think it might be a machine at the higher level of the hierarchy, which could solve the halting problem and then was worrying about harder things?

1:09:15 **David Deutsch** Yes, although calling it higher level, that's from our point of view.

1:09:20 **Charles Bennett** I mean in a Kleene hierarchy.

1:09:23 **David Deutsch** Yeah, which is, again, expressed from our point of view. It could just have different fundamental states and different fundamental operations. So one of its fundamental operations might be to solve the halting problem. Not very slowly, like you're envisaging, but instantly. So you could ask it questions about Turing machines and it could answer all such questions instantly. On the other hand, adding two and two would have it scratching its head for a million years.

Well, you can't deny that in the set of all mathematical objects, such things exist. And therefore it is logically possible that physics conforms to that object rather than the objects we think it conforms to.

1:10:41 **Charles Bennett** Okay, I admit that. I'm going to just disagree with you in a way that we can't prove very easily because Turing is dead, that he would have sympathized more with my view that all the different parts of physics, including the parts of the universe we can't get to because they're beyond the Hubble distance, would be simulable by a Turing machine and wouldn't involve these higher-level things. But seriously physicists, remember Hartle? Gerlach and Hartle wrote a paper about what an uncomputable number [would] look like if it was a physical constant. And I think that's certainly a legitimate question. And you're saying that those things, that one aspect of the mathematician's dysphemism is that they assume that such things don't exist.

1:11:43 **David Deutsch** Yes, yes, when applied to physics, yes.

1:11:48 **Charles Bédard** And I think David was also invoking not just beyond the Hubble distance, but also in a completely different universe.

1:11:58 **Charles Bennett** Yeah, well, that's what string theory gives us as far as I understand, Andreas, that there are things where I guess it's consistent with quantum field theory [and] general relativity, but things that are just extremely different from anything that we easily imagine.

1:12:14 **Andreas Albrecht** You don't even need string theory. I would say the development of theoretical physics keeps pulling us in that direction and it's hard to truncate. I don't know if it's right.

1:12:32 **Charles Bennett** Yeah, even uncomputable things could be there?

1:12:38 **Andreas Albrecht** I don't think that way, so I don't know how to answer that. But I think one of the fascinating things about cosmology is that physicists pride ourselves in this culture of—we only talk about real stuff we can touch and measure, but cosmology really disrupts that. It's really hard to write down theories with the laws we have, with the tools we have that allow us to limit ourselves that way.

And I'm actually really intrigued by David's angle, which seems to be to take that as a judgment of our ideas about physical laws, if I understand that, which I find fascinating. I've come at a lot of these ideas from the point of view that physics has nothing to do with infinity because the only stuff you have to work with is finite in terms of having finite data and so on. And what I'm hearing, very directly from David and I think implicitly, or maybe directly too from Charlie. Anyway, I'm hearing in both your comments that the problem of the arrow of time or the problem of Boltzmann brains or however you wanna put it, is looming enough to transcend. That's forcing us away from that position. And I will say that wherever I am in my prejudices and all of that, I think that that problem looms so mightily that I have to respect, even though you're doing stuff that I find uncomfortable, that problem looms so mightily that I respect.

It's very radical. To me, it's very radical what you're trying to do, but something has to give.

1:14:56 **David Deutsch** Oscar Wilde said, "We're all in the gutter, but some of us are looking at the stars." Well, some of us are looking at infinity because it solves problems here and now.

1:15:11 **Andreas Albrecht** But it seems to me like a cheat. I think you're also giving up. And I think you're explicitly saying that you're giving up on the standard ways we think about doing physics. Whether it's embracing some beautiful, better thing or giving up is a little hard for me to tell right now.

1:15:33 **David Deutsch** We've embraced the continuum for centuries.

1:15:40 **Andreas Albrecht** But that's only to make our life easier. And if it were discrete instead, it wouldn't change it. It wouldn't be a radical thing to put it on a lattice.

1:15:54 **David Deutsch** We're not here to make our lives easier. We're here to understand the world. And that's why these continuums and derivatives and all that infinite stuff was invented, was conjectured. And it might be false, but it might not be. We shouldn't have prejudices about these things.

1:16:12 **Andreas Albrecht** Yes, that's fine. I like your line. I find the finiteness of physics, physics is finite. What I'm hearing is, "Okay, physics might be finite, but except for the arrow of time problem forcing us out." I don't think the continuum forces us to infinity. We can have a lattice, we can have all kinds of things that do just fine, but you're saying this is the one thing that forces us to think about infinity, which is intriguing.

1:16:52 **Charles Bédard** I'd like to go in the audience and give the chance for some students to ask questions, if any one of you have a question.

1:17:03 **Audience Questioner** Hi, I'm not sure whether I understood why incompleteness is fundamental for non-equilibrium phenomena and so forth. I think it was a point of David that he made before. So I'd like maybe more comments about that. Thank you.

1:17:26 **Charles Bedard** Cool, thanks. David, would you like to react?

1:17:30 **David Deutsch** I'm sure Charlie could do this better, but the connection is that…you have to look at the contrapositive. If there is a limitation such as Hilbert imagined, that you can write out an algorithm that will be a criterion of truth, then mathematics has stopped at that point. And the same would be true of physics. If there was a knowable law of physics that predicted everything, then that would be the end of physics. And if there were such a thing as life in a universe, which I think there couldn't be, but if there were, then it would really mean that the design of the most complicated creature that could exist, which would be finitely complicated, would exist baked into the laws of physics at the Big Bang. So the universe would not have any of the kinds of openness to explain the complexity and all the high-level structure if there were infallible truths available. [To] put that the other way around, unless there were fallibility.

1:19:21 **Charles Bennett** Yeah, I think that's right. It's because incompleteness is the flip side of universality. If you restricted the laws of nature to something in which you couldn't produce a universal computer, of course, it would be limited by how long it could run before, but basically something that was behaving like a Turing machine until some part broke, then you couldn't get something as complicated as a bacterium.

1:20:13 **Charles Bédard** Thanks. Are there any other questions from the audience? Yeah. Vincent, would you like to comment?

1:20:20 **Vincent** So, hello. Thank you so much for the great discussion. Maybe I have a question about a bit of a different topic, namely about artificial intelligence and the AGI maybe, because some of us are interested in this. So from the Turing-Church thesis, we know that our brains are not doing anything magical. They could be simulated by a Turing machine. So that means, in principle, we could implement something like AGI if we would know how. But it could be that the only way, at least in my view, the only way to implement this is to simulate something like a human brain on a Turing machine. Not that there's any simple AGI program that does something like human-level intelligence or creativity, or maybe there is any reason we might get there. Do you think there is any reason that we should expect a simple program that implements AGI, or do we have to take the detour to simulate something like a human brain and this then?

1:21:32 **David Deutsch** Well, the human brain might be simple [if] looked at the right way. I think 'simple' doesn't necessarily mean easy to find. I think there is a very strong reason to believe that the explanatory universality property of the human brain, which at present only humans have, must be encoded in a very short amount of DNA. A tiny amount. People have different numbers for this, but we seem to be 95 percent or 99 percent of our DNA is the same as chimpanzees. And of the rest, a lot is junk DNA. And of the rest, a lot is differences between us and chimpanzees that aren't connected with these deep epistemological things. So maybe it's only a few K of code in DNA terms that encodes the AGI-ness or the GI-ness of the brain. But that doesn't mean it's easy to find. I would guess that it is a short program. A few K of program is actually long if you want to write it, but it's relatively short compared with the kinds of things that happen in biology. And I think it'll be very hard to find. And probably

I would expect us to find it only once we have understood a philosophical theory of what the function is, what the creativity, consciousness, qualia, and so on, what those things are in some precise terms. Precise, but high-level, I would expect. And then we can probably quite easily write a program that has that property. Then we'll be in various kinds of trouble, but not any of the kinds of trouble that people are thinking about now.

1:23:44 **Charles Bennett** I worry that people who know more about it are scared about it. But people are so good at being cruel to each other with their intelligence that I doubt that the machines would do much worse. And I take this Schopenhauer's principle, which is what they should call the anthropic principle, pretty seriously. And it may be that we're not likely to last more than a few more decades. I don't think human civilization will extinct itself, but it might set us back a few centuries or millennia. And we just hope for the best. So that's why I don't worry about it so much, but maybe I should. It's also, if you want to be universal, not in the sense of computational universality, but in the sense of the Universal Declaration of Human Rights—where do we get off thinking that we have better rights than these machines? They might actually do a better job, or they might be our successors, not in the sense of the ones that conquered us and enslaved us, but in a way that provided a shortcut to the bad instincts, or let's say the maladaptive instincts that humans have, and that it might take a bit of very unpleasant natural selection to get rid of. And maybe the artificial intelligence would make a shorter shortcut to...

1:26:09 **Andreas Albrecht** So you're saying AI might be better people.

1:26:12 **Charles Bennett** Yeah, yeah. In other words, another way of doing it would be giving us a transplant of bonobo genes or something that just makes us more susceptible to taking it easy and less susceptible to getting angry at our neighbors because of misinformation we've heard about them and going out and killing them, which seems to be the tendency people have, which probably was very adaptive at one stage, but it's not helping us right now.

1:26:47 **David Deutsch** Thanks. You don't believe in the better angels of our nature and Steven Pinker and all that stuff, he says we've been getting better and better.

1:27:00 **Charles Bennett** That [is a] tendency, but there are also strong tendencies in the other direction. We saw in the twentieth century, it got pretty bad and it could do so again. And it's worse now because the reach of things is even more global than it was in the twentieth century. So a very repressive world government would be harder to dislodge.

1:27:29 **David Deutsch** There's a more fundamental reason why you're right, namely because of fallibilism. There can be no upper bound to the size of error we can make. So although we have the potential to go exponentially into the future ad infinitum, we have the potential to make arbitrarily large errors. We may have a dark age and another dark age and a dark age lasting a million years, or we might wipe ourselves out and something else might evolve somewhere else in the universe or somewhere else in the multiverse. There can be no guarantee that that won't happen. There can be no guarantee that it probably won't happen. So what we have to do is solve problems as we find them and create knowledge as we can and not rely on supernatural guarantees. I thought you'd like that point about the finite bound.

1:28:41 **Charles Bedard** Are there any other questions from the audience? Aditya? I'm changing chairs depending on which side people come.

1:28:58 **Aditya** Hi. So I had a question because a lot of this discussion we spent discussing about how undecidability and the halting problem are, I guess, flip sides of the same problem. Universality and the halting problem are the kind of opposite sides of the same problem. Would you also say that these are opposite sides of induction and deduction as ways to kind of acquire knowledge? Because we talked about Solomonoff induction a little bit in the beginning and how maybe that is one conceptualization of how you could derive knowledge, instead of using deduction from a formal set of axioms where you would use data instead. So would you say that the incomputability of something like ideal induction is the same as not being able to deduce because of incompleteness?

1:29:57 **David Deutsch** Yes. They are both impossible for the same reason and their impossibility is a very good thing for the same reason. So in both cases, instead of deriving things, we have to guess things. We have to conjecture, and the conjectures are always fallible. And I'm spouting Popper's philosophy here. So that's where you have to go for this. And he got an amazing number of things right working almost in isolation.

1:30:34 **Charles Bédard** To relate to Aditya's point, I think also [Solomonoff] induction is an idealization, and to have Solomonoff's metric, one needs to solve the halting problem. But in concrete applications, one cannot find Solomonoff's prior. And so what we do, we can upper semi-compute it perhaps, and then come up with guesses, "Oh, maybe this phenomenon from which I got these data has been explained by this program." [Now] everything is programs, but one can come up with this. Actually, that's also a problem: how we link scientific explanations with programs. But I think there is a bit of a Popperian flavor in Solomonoff induction once we realize that it's uncomputable. So forget it. It's a beautiful idea. It does converge to whatever it needs to converge. But the fact that it's uncomputable, we can only guess programs.

1:31:32 **David Deutsch** Often the problem we're trying to solve isn't in data. We haven't yet got any data. Sometimes it's a theoretical problem, like, "How is it possible for Maxwell's equations to be true and geometry to be what we think it is?" There's no data, no data at all. So you have this problem first, then the theory, then the data. And so induction of any kind simply can't exist in that kind of a reality. And as I keep saying, it's a very good thing that it can't. So Solomonoff is trying to solve a problem that isn't there. It's trying to say, "How can we make this induction or this Bayesian inference or whatever makes sense? How can we get the priors right?" Well, the priors aren't right. Get over it.

1:32:39 **Charles Bennett** Well, my colleague John Smolin told me about the story of induction and about an explorer who comes to a place that's inhabited by anti-inductionists. And these are people who believe that if something happened once, it's less likely to happen again. And they have a miserable life because their buildings fall down and their crops fail. And so this explorer says, "You know, I understand why everything is not working here. It's because your principle of anti-induction is wrong." And they looked at him and they said, "Why should we give up this principle? It's never worked before."

1:33:32 **David Deutsch** Nice.

1:33:37 **Audience Questioner** Thank you. Thank you so much for your answers.

1:33:43 **Charles Bédard** Other questions? Sam?
1:33:49 **am Kuypers** I'm interested in what kinds of cosmology Charlie is interested in. We have the cosmology for people like Tipler. I'm just curious what interests him at the moment.

1:34:09 **Charles Bennett** Well, I don't know a lot about it, but I've been trying to understand these models they call eternal inflation, where there isn't a beginning or an end, but in which there's inflating pockets that appear here and there with all sorts of different natural laws in them. And in that, that's a pretty scary place to live. And it's populated by all kinds of problems, like David mentioned. It's hard to get a universe in which you're not extremely likely to disappear in a microsecond from now.

1:34:59 **Charles Bédard** Other questions from the group? If they come up later, feel free to raise your hand at some point. And at this stage, I could also take questions from the online event.

1:35:16 **Sophie** I would have a quick question.

1:35:19 **Charles Bedard** Go ahead, Sophie.

1:35:21 **Sophie** I'm wondering when you're talking about other universes that they could solve the halting problem, and for them it's pretty easy, but for us it's difficult, and vice versa for other tasks. Are you thinking about other branches in the multiverse or something that would be completely outside of what we are studying now?

1:35:49 **David Deutsch** The latter. All the universes in the quantum multiverse have the same tame mathematics. They even have the same computable functions. They just have different complexity theory, but complexity theory isn't that important, anyway. We have to consider universes with different laws, if only to examine explanations like the anthropic principle, which purport to explain things in our universe and only certain very special kinds of anthropic argument constitute arguments. Otherwise they simply fall victim to this Boltzmann brain or boundary of the set of possibilities problem. Dennis Sciama, when he told us about this objection, he said that what we should really look for, anthropically, is not a case where a slight change in a fundamental law would have meant that we're not here, but where we are located at the center of the region, let's say, the fine structure constant, rather than saying, if it was one percent different, we wouldn't be here. Look at the region where we would be here and try to find out whether we're near the center. If we're near the center, that is a real anthropic problem. If you tell this to religious people, they're going to say it's God. And it's very hard to argue them out of it. But I don't think we will be in the center.

1:37:51 **Charles Bédard** But I think you don't think we're also on the edges, because if we're on the edges, we're out of it right after.

1:37:58 **David Deutsch** Yeah, we're neither at the center or the edges, because that's not where the explanation lies. The explanation of the actual values doesn't lie in our existence. It's a common cause. Our existence is caused by things like the law of the universality of computation.

1:38:33 **Charles Bennett** We're not at the middle and we're not at the edge. Exactly how far from the edge are we?

1:38:39 **David Deutsch** Probably a non-computable amount or an intractably computable amount.

1:38:51 **Andreas Albrecht** David, I'm just curious, this is a small question, but I'm just curious. So your comments about the sort of unpredictability of the future of our civilization and the possibility of our demise, is that the same thing as this edge that we've been talking about? Or do you see a difference?

1:39:13 **David Deutsch** Yeah, it's almost the opposite. It's to do with our fallibility, which is to do with our capacity for infinite growth.

1:39:25 **Andreas Albrecht** Good, that's interesting. So it's our hope as well as our possible demise?

1:39:34 **David Deutsch** Well, I'm not expecting our possible demise. I think arguing that there will be a demise is itself not legitimate. That's a prophecy. That's a prediction.

1:39:49 **Andreas Albrecht** But it's in the scope. You're saying we can't be sure to avoid it.

1:39:58 **David Deutsch** The future of knowledge is unknowable. Again, we've got to get over that. It's a fact.

1:40:06 **Charles Bennett** You're saying unlike other predictions, it's not false survival. Because if there is a demise, it's like Gilles Brassard wrote a book on theory of computation, in which it had a cartoon of a bunch of elderly scientists looking into a steaming test tube. And one of them is saying [that] we may have discovered the elixir of immortality, but it will take forever to test it.

1:40:48 **Charles Bédard** Any more questions?

1:40:57 **Sophie** I would have another question maybe for Charles. I wonder what's your view of probabilities?

1:41:07 **Charles Bennett** My view of probability? I think it arises from entanglement.

1:41:27 **David Deutsch** Yes. Agreed.

1:41:28 **Charles Bennett** Probabilism is a feature of a subsystem in a larger system that evolves unitarily.

1:41:45 **Sophie** Cool. My question could also be formulated as: Do you agree with David in terms of what are probabilities? And I guess the answer is 'yes.'

1:41:55 **Charles Bédard** Not necessarily, because many people view probabilities as entanglement, because they understand quantum theory as unitary. And so entanglement is out there. But even among these people, there's still many disagreements about how we should read the Born rule. And so maybe the follow-up question for Charlie, without being so specific as does it equate to David's, despite you embrace unitary quantum theory still, how do you view the Born rule? Is it a decision?

1:42:27 **Charles Bennett** Oh yeah, it's hard. I try not to think about it, but I know how to. David, do you have an answer on that?

1:42:45 **David Deutsch** Yes, I think the decision theory approach to probability in quantum mechanics is correct and sufficient. And whether you call that deriving the Born rule, I don't know. I think the Born rule is not a universal rule, obviously, because it only applies to quantum theory in a tiny minority of situations. The decision theoretical approach is to take classical decision theory and quantum theory, strip out all the probability from classical decision theory. So strip out all the probabilistic axioms and so on. And then quantum theory, strip out the probabilistic, if you're going to have wavefunction collapse or something, strip out all that stuff, anything that's supposed to be probabilistic. Then smush them together and you get consistent quantum theory, namely Everett's, and you get a consistent theory of probability, which only applies when people are making decisions. And the rule that they use, assuming that they make measurements accurately and all that stuff in that approximation, their rational behavior as determined by decision theory is the same as it would be if they believed in stochastic processes, even though stochastic processes do not happen in nature.

1:44:40 **Charles Bennett** Yeah, that sounds pretty plausible.

1:44:46 **Jean-Michel Lemay** Cool. I'm not sure I properly understood your explanation, David, because if you say you want to get rid of the Born rule and you believe in a unitary quantum mechanics and Everett's many-worlds interpretation, how do you reconcile the fact that when we do a quantum experiment in the lab, we just get to see one result? How do we interpret the fact that we're in just a specific branch of the multiverse and not another one?

1:45:20 **David Deutsch** We're not, we're in all of them. And the fact that we see only one outcome is a prediction of quantum theory. If you just write the equation with the variable for how many outcomes do we see, then you will find that the world after a good measurement, the world will be in an eigenstate of that with eigenvalue one. What actually happens is that all of the outcomes happen, generically. But the question was about probability. What do we then mean that some of them have a high prior probability than others, given that they all happen? Then you have to bring in another part of the model, which is: Suppose we were betting on these outcomes, how would it be rational to bet? And decision theory without probabilistic axioms and quantum theory without them together give the answer. It's magic. I'm not surprised you're shaking your head, but if you read the papers, it's undeniable.

1:46:40 **Jean-Michel Lemay** Do you think it's unanswerable, or we haven't found a good explanation for it yet?

1:46:46 **David Deutsch** 'Unanswerable' is too ambiguous a word. It's undeniable that this is the explanation.

1:46:54 **Charles Bennett** So David has, amidst his other euphemisms and dysphemisms, brought in what is, I regard the most widespread obfuscatory euphemism: God. People use it to mean their own idea of God, which they're going to try to get you to like and believe in, but they haven't actually said what it is yet. So I'm going to use that also, but in a physicist's way. So if you imagine a conversation with God, and first of all you say, "What time is it?" And God says, "Well you would say it's about 10:37 Eastern daylight time. But for me, it's all times, I'm eternal." And then you'll say, "Well, you're supposed to be omniscient. Will it rain a year from now?" And God says, "Well, yes and no." And you'll say, "Well, will it rain for me [a year] from now?" Well, it's like saying, "What I see is something that looks like your hand and you're sitting here and you're saying, 'Which of my fingers is the correct extension of my wrist?' They all are. When it gets to be next year, you'll either be on one where it's raining, which is the one that has my wedding ring, or you'll be one of the other ones where it isn't raining. And I can see all of them but, for you now, they're all equally real. And then next year, you will have a kind of a truncated perspective in which you think one of them is real and the others aren't, but they all still are."

1:49:07 **David Deutsch** I'm not going to think that. I know that the other branches exist even if I can't see them. Most things in the universe I can't see. I don't know why that's so…

1:49:21 **Charles Bennett** So this is like this. What Galileo said, I don't know if he actually said this, but he says, "Yes, I say that the Earth moves, and if it moves the way I say it is, you shouldn't be able to feel it moving so shut up."

1:49:38 **Charles Bédard** Yeah. Pretty much.

1:49:40 **David Deutsch** Everett said that to DeWitt.

1:49:42 **Charles Bennett** Yes. So did he use the Galileo example?

1:49:46 **David Deutsch** Yes, absolutely.

1:49:47 **Charles Bennett** Yeah, yeah that's right.

1:49:49 **Charles Bedard** And is that the moment where DeWitt started to say, “Okay I don’t know what more to argue against.”

1:49:54 **David Deutsch** Yeah. That’s the moment he was convinced.

1:49:56 **Charles Bédard** Yeah. So the idea is—even the theory says you will only experience one outcome. So even though we seem to be bugged by the fact that, “No, but the theory says there’s going to be many outcomes.” No, no, no it might look like the theory says there’s going to be many outcomes, but if you ask yourself, “How many outcomes am I going to see?” the theory says, “Systematically, only one.” So it preserves its consistency in spite of its striking apparent inconsistency. Are there other questions?

1:50:48 **Audience Questioner** Hey, do you hear me? So would you say or any of you say that in the sense we just discussed now, the Born principle would be a statement about consciousness of what do you experience? Or do you think it’s beyond, like consciousness would be beyond the scope of the statement?

1:51:07 **David Deutsch** It has nothing to do with it. Obviously if consciousness is to be described as a physical process, which it must be, then it presumably obeys quantum theory and so on, but it has nothing to do with these issues of measurement and probability and anything like that. It’s just a physical thing. There’s no more reason to assume it has a special place in physics than it would be to assume that squirrels have a special place in physics. Obviously they’re described by physics, but they don’t have any relevance to fundamental discussions of physics.

1:51:54 **Audience Questioner** I asked before because the idea was, if I’m following the discussion, is that you have these different potential outcomes of measurement, all of them occur, all of them have their own branches, and all of them are there. But the statement is more like, "You will experience one of them," and then if it's about you will experience one of them, you will see one of them, wouldn't it be addressing directly the consciousness, or maybe I'm just going out of it?

1:52:22 **David Deutsch** The same question can be asked about a seed of a plant that grows into two trees, the left one and the right one. And you ask the seed, "Are you going to be the left one or the right one?" Or twins: "Are you going to be Joe or Fred?" You ask the fertilized egg and it says to you, "Well, in a sense, I'm going to be both of them. In a sense, I'm only going to be one of them. That is to say, I will not experience being both of them, but the experience of being each of them will both happen."

1:53:02 **Audience Questioner** Okay, I understand.

1:53:05 **Charles Bennett** That's a good way of putting it. Yeah.

1:53:09 **Charles Bédard** Well, David and Charlie have been tremendously generous of their time. It doesn't [seem] like there's other questions. If somebody has a burning desire and has been repressing his question, this is the last call. [It] also applies for the audience. Great. So thank you so much David and Charlie. That was a lot of fun to discuss with you and it was very, very fascinating in many regards.

1:53:42 **Various** Thank you very much. Very interesting conversation. Thank you. Bye, everyone. Thanks a lot. This was great. Bye.

SEAN CARROLL:

SCIENCE, COMPLEXITY, AND EXPLANATION

About the interviewer: Sean Carroll is interested in how the world works at the deepest levels, which leads him to do research in physics and philosophy. His current interests include foundational questions in quantum mechanics, spacetime, statistical mechanics, complexity, and cosmology, with occasional dabblings elsewhere. His new book is *The Biggest Ideas in the Universe: Quanta and Fields*, where he introduces modern physics with all the details, in a way accessible to anyone. He hosts a podcast, *Mindscape*, where he interviews smart people about all sorts of interesting ideas.

His official title is Homewood Professor of Natural Philosophy at Johns Hopkins, and he is also Fractal Faculty at the Santa Fe Institute. He lives in Baltimore with his wife, writer Jennifer Ouellette, and two cats, Ariel and Caliban.

EPISODE DETAILS

Date	October 16, 2023
Interviewer	Sean Carroll
Source	YouTube
Show	Mindscape
Episode	Mindscape 253 | David Deutsch on Science, Complexity, and Explanation
Description	David Deutsch is one of the most creative scientific thinkers working today, who has as a goal to understand and explain the natural world as best we can. He was a pioneer in quantum computing, and has long been an advocate of the Everett interpretation of quantum theory. He is also the inventor of constructor theory, a new way of conceptualizing physics and science more broadly. But he also has a strong interest in philosophy and epistemology, championing a Popperian explanation-based approach over a rival Bayesian epistemology. We talk about all of these things and more, including his recent work on the Popper-Miller theorem, which specifies limitations on inductive approaches to knowledge and probability.
Link	https://www.youtube.com/watch?v=ldgK7EhEnto

Ideas

- We will soon be in a position where Earth will soon be in a position where if asteroids or comets head towards it, they will be repelled rather than attracted. For all we know, every other planet in the universe attracts asteroids and comets, and ours will be the only one to repel. That's a physical fact for which the non-anthropocentric explanation is that there is explanatory knowledge on the Earth.
- What's needed to create AGI is a new explanatory theory. It'll be largely a philosophical theory. It'll be a new way of looking at what creativity is, what explanation is. The program for AGI could be very simple in terms of number of bits. We differ from the great apes only by a few K of code. In that few K of code is the program for bootstrapping this qualitatively different type of program that we run as universal explainers.
- Popper taught us that the content of a scientific theory is in what it rules out. And if you just took that seriously as the basis of your worldview, you'd immediately come to constructor theory because then you would say: What does this given theory rule out, and what doesn't it rule out? And the distinction between those two is the theory. Popper never said that. If he had said that, then he could have discovered constructor theory.

Topics

artificial general intelligence • constructor theory • credence • Darwin • DeWitt • Duhem and Quine • Everett • explanatory knowledge • explanatory universality • fundamentality of knowledge • general relativity • good explanations • Hempel • Heisenberg picture • history of many-worlds theory • Kuypers • Marletto • Newton • obedience • paradox of the intransitivity of support • Popper-Miller theorem • Popperian epistemology vs. Bayesian epistemology • principle of optimism • probability • probability calculus • proposition • quantum field theory • quantum mechanics • the Enlightenment • the Great Monotony and explosion of novelty • Turing completeness • universal constructor • von Neumann • Wheeler

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Sean Carroll** Hello everyone, welcome to the Mindscape Podcast. I'm your host, Sean Carroll. We've talked about quantum mechanics a lot on the podcast. You may have heard that I'm a fan of the Everettian many-worlds formulation of quantum mechanics. We have a special treat in that we have a guest who actually met Hugh Everett and was influenced by him and has gone on to be a major proponent of the Everettian version of quantum mechanics. That would of course be David Deutsch.

And despite the fact that David is very well-known in his work in quantum mechanics and quantum field theory, basically if you have to give credit to one person for pioneering the idea of quantum computers, it would have to be David Deutsch. There's other people who made very significant contributions there, but David was one of the first to really define what it means to do a quantum computation, to write down an algorithm that was faster than a classical algorithm, to really think about how entanglement can help you encrypt things using quantum mechanics and so on. It's been super-duper influential. He's been awarded various prizes for this: the Breakthrough Prize, the Fellowship of the Royal Society, and so forth. But that's not all.

In fact, in this podcast, we're not even going to talk about quantum mechanics that much. We're going to be talking about various things that David has been thinking about that grow out of, arguably, his combination of [an] interest in the fundamental laws of physics, but also in epistemology—how we learn things about the world. You've heard me talk about quantum mechanics and Everett, you've also maybe heard me talk about Bayesian reasoning and Bayesian inference and epistemology. And so unlike quantum mechanics, where David and I are very much on the same team, here we are not. And so that's what I wanted to talk about. He's been thinking a lot about, I guess what you might call Popperian epistemology, after Karl Popper—the idea that we think about possible worlds and we divide them into the ones that are compatible with the data and then not, and then seek the best explanation. It's a little bit fuzzy, I gotta say, what counts as the best explanation, but it's clearly also very similar to what we actually do. I mean, you can recognize this in the actual progress of science. We try to come up with the best explanation for what the world is doing given the data we currently have and a way to go beyond that. So David has been trying to formalize that, thinking about it very carefully, and pointing out where traditional mottos that one invokes in the Bayesian context might be hiding some subtleties that make them less applicable than you might think. In particular, there is a theorem due to Karl Popper and Miller, I don't know what Miller's first name was, but the Popper-Miller theorem that David has been thinking about that he would argue, and I think there's a case to be made, makes it hard to accept traditional Bayesian vocabulary as how we really go about picking our theories. So that's a very interesting conversation to have.

And another thing that David has been interested in is constructor theory. I don't know if you listened to the podcast we did a while ago with Chiara Marletto, who is David's collaborator in this. They've been developing literally an entirely new way to think about what it means to do physics, to be a law of physics. Rather than having some dynamical law where you start with initial conditions and just chug forward, they think about physics—and not just physics but also biology, chemistry, et cetera—in terms of what is possible, what is not possible, and what kind of constructors can actually make things happen in the world. I don't know, I still don't know, after talking to Chiara and now to David, I still don't know whether this is going to be super-duper useful going forward. It might very well be, though. I'm very open to that. I'm very interested in seeing where that goes. So we talk about that, too.

We talk about the space of possibilities and how knowledge and explanation have burst onto the scene in the universe with the advent of human beings and their brains, and he's very careful to say it's not just necessarily human beings. Aliens, computers, could also qualify. But it's a dramatic shift in how the universe evolves when you have systems that can think, store information, come up with explanations, use that knowledge to transform the world around them. It's ultimately an optimistic perspective on the world, and that's something we could all use a little bit more of, so I think this is going to be a fun conversation.

Occasional reminders that we have a Patreon page here at the Mindscape Podcast. You can go to patreon.com/seanmcarroll. Kick in a dollar or two per episode, and the benefits will just start flowing your way. The benefits are not huge, but [there are] still benefits. You get an ad-free version of the podcast, you get to ask AMA questions, you get to participate in those discussions, and after every podcast I do a little reflection video and audio that is for Patreon members only. So join the fun there, patreon.com/seanmcarroll. With that, let's go.

5:39 **Sean Carroll** David Deutsch, welcome to the Mindscape Podcast.

5:42 **David Deutsch** Hi, thanks for inviting me.

5:44 **Sean Carroll** We're gonna get into substantive stuff soon enough, but I've gotta start with a question I've had for a long time. I believe that you were in the audience for a seminar given by Hugh Everett at the University of Texas some time back. Is that true?

5:59 **David Deutsch** Indeed, I was.

6:01 **Sean Carroll** Can you say, was it actually kind of a formative experience? What was it like? What was Hugh Everett like?

6:06 **David Deutsch** It was a memorable experience. I had imagined him differently. And I knew that Wheeler invited him. I was quite excited that he'd invited him because very few people were Everettians at the time. I suppose very few are now still.

6:33 **Sean Carroll** This is the seventies?

6:36 **David Deutsch** Yes, in the seventies, the late seventies. Wheeler had invited him and was treating him like royalty. And one example I remember, I can't remember exact details, but one example I remember is that there was [a] strict 'no smoking' rule in the seminar room. And that was quite rare in those days. It hadn't yet become ubiquitous like it is now. But Wheeler asked for this to be waived in the case of Everett because he was a chain smoker. He didn't stop smoking. And so this leaning over backwards to make him feel comfortable. And he gave a talk about the Everett interpretation, or Everettian quantum theory as we now prefer to call it. And then we went to have lunch because the graduate students and the postdocs and the faculty on our floor often used to go and have lunch in one of the places in Austin. And Bryce DeWitt contrived to have me sit next to Everett. So I had lunch chatting to Everett, and I asked him some elementary questions. I hadn't really started thinking very seriously about it, and I was just very impressed that he was completely on the ball, up to date with all the nuances. And we had a nice chat, and that was the last I saw of him.

8:26 **Sean Carroll** Well, that was my impression that I only got from finally writing a book about quantum mechanics. He only wrote the one paper, and we've all known physicists or scientists who have the one paper and they move on and maybe they were [in the] right place at the right time, but you get the impression that he really did very much understand all the nuances that were going on. It wasn't just that he got lucky.

8:49 **David Deutsch** Yeah, very much so. And I got the impression, although I know other people have a different history in mind, but I got the impression that he did not leave research in physics because he was disappointed at the reception his ideas got or anything like that. He left it because he wanted to do other things and to make a fortune and, you know, he did.

9:15 **Sean Carroll** You know, that's perfectly valid, and I also got that impression. It sounds like you were Everett-sympathetic even before that talk, but did that inspire you to think more carefully about it?

9:26 **David Deutsch** Yes, I was Everett-sympathetic because of DeWitt. So I was lucky in that respect. I met DeWitt and then a couple of years, I can't remember how long later, Everett. So I met DeWitt when DeWitt was on sabbatical in Oxford. I was a graduate student. I was a first-year graduate student. There again, not by anyone's contriving, but by sheer chance, I was in Dennis Sciama's department and we went to have lunch at a pizza place in Little Clarendon Street in Oxford, and I happened to be sitting opposite Dewitt and I only vaguely remembered that Dewitt had something to do with the Everett interpretation. So I thought, well, I'll ask him, and I asked him a very silly question. I can't remember exactly what it was, something like, "If there are many copies of me, which am I?" or something like that, something as elementary as that. And he was very kind and explained to me that this was not a good question, and the way to think about this, and he explained to me. Then I asked him more questions. By the time we'd finished lunch, I was completely convinced. I mean, previously, I'd already thought this was worth looking into, because it was a version of quantum theory that was purely physics and didn't have any kind of psychology or assumptions about the brain and that kind of thing. So it's just how we'd been taught to deal with theories. But what convinced me was that lunch with DeWitt.

11:25 **Sean Carroll** It actually leads into the broader conversation because you're giving examples of how the space of possibilities in life is very, very large, and tiny, unexpected events can steer you in one direction or another.

11:39 **David Deutsch** Well, yes. I think that's true, but I'm not sure that these examples were examples of it because I'm not sure, well, I'd like to think anyway, that I wasn't exactly steered. I was just hastened. I think I would have come [around] to this eventually, if for no other reason than I would eventually have read DeWitt's work on this and Everett's, and I would have talked to Wheeler about it and been dissatisfied with his answer. So, you know, I think that would have happened.

12:17 **Sean Carroll** There's some convergent evolution there, yeah.

12:19 **David Deutsch** Yeah, exactly.

12:20 **Sean Carroll** Good. We've already mentioned that there's a lot of things to talk about, but I've chosen as the substantive starting point—monotony. You gave a nice little TED Talk on the end of monotony and how we're moving into a different era. I thought that the title of the talk was maybe not the most inspiring, and maybe you would get more clicks if it were not about monotony, but maybe you could explain what the basic idea there is.

12:48 **David Deutsch** As always, the titles are not chosen by the author. That title was not chosen by me. So yes, it seems that progress is not uniformly rapid. And progress in various senses, like the origin of planets like ours, and the origin of life like ours, and the origin of multicellular life, and the origin of explanatory creativity as in humans, and then the origin of the explosion of the Enlightenment. All those things happened very rapidly after a long period of not happening. In all cases, you can't really put your finger on why it took so long. I think in a couple of cases, we'd say it's not surprising that it took so long because it was rather a big step, but why did it take billions of years in one case? Why did it take thousands of years in the case of the Enlightenment? We don't know, but it appears it happens that way. By the way, in case you're going to ask, this is not punctuated equilibrium. This is not a substantive theory about how or why adaptations or knowledge happens. It's not that there's an equilibrium. I think in none of these cases was there an equilibrium. All of them were unstable to this thing happening. So there wasn't an equilibrium and the punctuation didn't have anything to do with how it then went on. It could have gone unstable in a different direction. So this punctuated equilibrium as advocated by Gould, for example, in my view is not a theory of evolution. It's just, at best, a description of what sometimes happens.

15:06 **Sean Carroll** Yeah, okay. All the words you're using about, you know, lasting a long time and then suddenly something happens, this sounds like phase transitions and metastability to me as a physicist.

15:19 **David Deutsch** Yes, so [the] difference between phase transitions and all this other stuff is that we can form a theory of when a phase transition is possible and then when it will happen. And if it's too complex to work out, then we can produce a better theory of it that predicts it better and [a] high-level theory and so on. So it's kind of [a] deterministic thing. And all the other things that I said are indeterministic things. They are things that some people would say [are] probabilistic, but I think that's not a good enough take on it either, because they're not something where the probability of it plays an important role in why it happened. If multicellular life had a probability of ten to the minus six or ten to the minus seven per unit time or something, neither of those explains anything. And if we were told that the probability of multicellular life evolving per unit time was actually one in a million years, we'd be wondering why it's a billion. But saying it's a million doesn't help to explain why it was a billion. You know, we'd need something else, something substantive.

16:57 **Sean Carroll** Are you pointing toward having a theory of why these things can bubble along, unchanging for a long time and then suddenly change gears?

17:05 **David Deutsch** No, I'm just being blindly critical of expecting everything to be known or expecting every regularity or irregularity in nature to have an obvious explanation. Some of them have an explanation, and when we find an explanation, that's great. I expect eventually we will find explanations for all these things. But I don't like this jumping into thinking we know almost everything now. We know almost nothing. During the pandemic, I was tweeting all the time, "This isn't known. Why are you writing as if this is known?" whatever it was. A lot of things were not known and a lot of things still aren't known.

18:03 **Sean Carroll** That's perfectly fair. Yeah. So you point in the talk to the origin of life as something that really changed things, that in a real sense for billions of years, the things that existed in the universe were the same things that existed a billion years prior, and something very, very new has come on the scene now.

18:23 **David Deutsch** Yes. The new thing is knowledge, and knowledge of particular kinds in all these cases. Some people will say, "Well, why is that particularly important? You know, knowledge is important to us humans because our ecological niche depends on creating and manipulating knowledge." But as I say in my book, if koalas could speak, then they might say that eucalyptus leaves are important and the emergence of eucalyptus leaves was... Now, I don't think that is so because knowledge is different from eucalyptus leaves, both from the point of view of understanding it and from the point of view of it affecting things. So in terms of it affecting things, an example that I like to cite is that we will soon be in a position where our planet or the planet Earth will soon be in a position where if asteroids or comets head towards it, they will be repelled rather than attracted. For all we know, every other planet in the universe attracts asteroids and comets, and ours will be the only one to repel. Now, I don't need to say anything anthropocentric to note that fact. That's a physical fact, and it's the same as the other kinds of physical facts where we say, "This phenomenon is different from that phenomenon," or we want to explain why. And the explanation in this case is that there is explanatory knowledge on the Earth.

20:18 **Sean Carroll** And I like the way that you put it—that in most cases in the universe, I'm going to paraphrase here, but big things push little things around, and knowledge has flipped that on its head in some sense.

20:32 **David Deutsch** Yes, so that is exactly the explanation for this purely physical thing that we have noticed. And then it's also true the other way around—that if we try to understand what's happening on Earth, then you will see, again, the example I give is that you will see people in the laboratories where they are looking for extraterrestrial intelligence, they will have a champagne bottle in the fridge, ready for an event that they are hoping for. And if you were an alien looking down on the Earth with [an] ultra high-[powered] telescope and you noticed that that champagne bottle was there and you wanted to predict something about that champagne bottle: Will it always be there? Will it stay there? When will the cork pop out? That sort of thing. You'd see that it's not just SETI. I gave the example of SETI, but really any team that is looking for a breakthrough might have such a champagne bottle in their fridge in their department. If you want to understand the behavior of those champagne bottles, you must understand not just humans, not just what happens on Earth. You must understand whether there's extraterrestrial life, whether quasars do this or that, dark matter, dark energy. You need to understand basically everything before you can understand how champagne bottles behave on the surface of the Earth. That again is because of the peculiar properties of explanatory knowledge.

22:29 **Sean Carroll** Maybe go into that a little bit more. I mean, certainly it is a feature of life, even in primitive organisms, that living organisms have some information about their environment and use that. They leverage it, right? And human beings do so in a more dramatic way. Are you pointing to the latter there?

22:53 **David Deutsch** Yes. So you might say this is only a quantitative effect, but, as Richard Dawkins says, every genome has got a blueprint of the environment that caused it. So the environment that caused bats or birds or something tells us something about, if you didn't know the Earth had an atmosphere, you might be able to infer it from the genome of bats or birds. But that's very parochial. The amount of the world that affects that genome is very tiny by cosmic standards, whereas the connection that I just mentioned goes all the way to quasars and to the Big Bang and to the end of the universe and so on. There's nothing in the physical world that can't affect those champagne bottles and only via the intercession of explanatory knowledge.

24:07 **Sean Carroll** And do you think that it's fair to attribute that to specifically humans here on Earth? I mean, there's going to be debate about what nonhuman species really understand.

24:17 **David Deutsch** Yes. So I prefer, when talking about these deep things, I prefer not to refer to humans specifically because if there are extraterrestrial civilizations, for example, then they will necessarily have this property, too. Because they couldn't have become civilizations and make flying saucers and so on without explanatory knowledge. And the same will be true once we have artificial general intelligence—they will also have [this property]. I prefer to talk about all those kinds of entities as people. Humans are people, extraterrestrials are people, AGIs will be people. And I argue in my book that there's nothing beyond that. Like, there may be AGIs that think many times faster than we do, but there aren't any that are in principle capable of connecting the universe with champagne bottles any more than we can.

25:28 **Sean Carroll** That's a crucial point. I wanted to get into that. So you think that we have crossed some threshold where things that are understandable, we can understand in some sense?

25:39 **David Deutsch** Yes, I think we have. And I think I have a watertight argument for that.

25:47 **Sean Carroll** So what is that?

25:49 **David Deutsch** It's in two parts. I think human brains have two kinds of universality that are essential to this. One of them is fairly uncontroversial among sort of scientifically minded people, and the other one is very controversial but I think just as compelling. The one that's uncontroversial is that human brains are Turing-complete. That is, we can execute any program that can be executed at all. Now, it might take us more than a lifetime. It might require more memory than we have, but we can augment our memory. We can augment our lifetime either by living longer or by having a tradition of doing certain things over generations. So those things aren't essential. We're accustomed to saying that the computers that we're having this conversation over are Turing-complete, even though they have only finite speed and finite memory capacity. But we know that those are trivial restrictions because they can, however complex the program that we want to execute with them, we could do it if we had a bit more memory and a bit more speed.

27:15 **Sean Carroll** Maybe for the audience, define what it means to be Turing-complete?

27:19 **David Deutsch** This was defined by Alan Turing in 1936 when he set up the modern theory of computation. He invented what we now think of as rather strange computers. They were made of paper tape and could move backwards and forwards via a reader. And he proved mathematically that a particular one of these could compute anything that any other one could. And this was a bit of pure mathematics. He also conjectured that the set of all of them was the set of all things that could be computed. In other words, that his model of computation was complete. Nothing could compute any more than that. He conjectured it, but once we went to quantum computation, I was able to prove that given quantum theory. So if quantum theory is false, it might still be false. But if quantum theory is true, then Turing's conjecture is now proven. So we know that there's only one kind of universal computation and that there's nothing beyond that.

28:37 **Sean Carroll** I think that maybe people have heard that before, but I think maybe it just hasn't made as much of an impression as it should. I think this is worth shouting from the rooftops, right? Like, not only can we calculate things and compute things, but we have very good reason to believe that even if we're slow and we make mistakes [and] whatever, the kinds of computations that can be done are kinds we can do.

29:01 **David Deutsch** Yes. We're as confident as we can be that when the aliens visit us or when the AGI become our new overlords, that they will not be able to compute non-Turing-computable functions. So that's as—or more—known to us than other bits of science or bits of physics. So that's the uncontroversial part. Although you say many people aren't, you know, it's not so familiar to many people, yes. By the way, Turing-completeness is a property of hardware. It's a property of the brain, it's a property of computers. The other kind of universality, explanatory universality, is a property of software, which I say we have. Our software has that property, and no other surviving organism on Earth has explanatory universality, although we know basically for sure that there used to be species related to us on Earth that also had explanatory universality, and they died out, which should be a warning to us.

30:32 **Sean Carroll** What do we have in mind there?

30:34 **David Deutsch** Well, like Neanderthals and I think going back to Homo erectus. Anything that had campfires necessarily has the thing that we have. Again, there aren't gradations of it. In the same way [that] there aren't gradations of Turing universality. You either have it or you don't. It's possible that you're rather impeded in using it because you don't have enough memory or whatever, but the basic thing is all or nothing. And I think the same thing is true of explanatory universality because, if I can put it in my idiosyncratic way, which I like, it's to do with optimism.

So the principle of optimism is that everything which is not forbidden by laws of physics is possible with enough knowledge. And the argument for that is that if there was something that was permitted by laws of physics, but could not be attained no matter what Turing-computable program we ran in our brain to do the thing, then it wouldn't be possible. And we could then test the scientific theory that that thing isn't possible after all. And that what we thought of were laws of physics were, in fact, not sufficient laws of physics. There would be, no matter how we tried, no matter what we tried, we wouldn't be able to do this thing, like exceeding the speed of light or whatever. It would be like that if it was building a certain tower or building a certain society. Either it's forbidden by the laws of physics or it's permitted, because if it weren't permitted, then you could do this experiment and, by the definition of science, you could set up a refutable theory and so on. So I think there's no getting around that. Therefore, I think that, just as there is only one kind of hardware universality, there's also only one kind of software universality, and that's the kind we have.

32:58 **Sean Carroll** Do we have a definition of explanatory universality that is as rigorous and mathematical as Turing completeness?

33:06 **David Deutsch** No, because there's a quite deep reason for that. It's because you can't formalize the notion of an explanation. You can always invent new modes of explanation, and they are conjectures like any theory. So you might conjecture that so-and-so is a good mode of explanation. The openness of science is connected with the non-formalizability of explanation. And by the way, that's exactly the same as the non-formalizability of mathematics. You can't formalize what is a valid proof because, however you formalize it, you can prove that there will be mathematical truths that can't be reached by that formalism.

34:02 **Sean Carroll** Is it then fair to say that, even if we don't have a rigorous mathematical definition of explanatory universality, we have a rigorous mathematical understanding that we never will have a rigorous mathematical definition?

34:16 **David Deutsch** Yes. Actually, interesting point. I never thought of that. Yes, I think we do.

34:21 **Sean Carroll** Okay, good, very good to know. But I want to sort of finish up the monotony discussion by reinforcing your optimism that you already mentioned. You make a good case, gently laying it out, that we're just at the beginning of truly transforming the universe based on this knowledge and explanatory power that human beings have developed.

34:48 **David Deutsch** Yes, I think that is necessarily true because the openness and the unboundedness are really the same thing. And again, the same thing is true of mathematics. We know that there's an infinite amount of mathematics to be discovered, even though, in the case of mathematics, there's a lot of it that we can't discover, unlike in the optimism case. I have a conjecture that we can discover all the interesting things, which are also infinitely meaningful.

35:28 **Sean Carroll** Oh, okay. Well, you have mentioned a couple times AGI, artificial general intelligence. I take it that you're relatively optimistic [that] that's on the way?

35:41 **David Deutsch** Depends what timescale you're talking about. I think we do not have the slightest clue how to make an AGI. I think what's standing between us and making an AGI is an explanatory theory. It'll be largely a philosophical theory rather than a computer science theory or mathematics or physics or anything like that. It'll be a new way of looking at what creativity is, what explanation is. And I think that, qua computation, qua computer program, I would expect it to be very simple, relatively simple. So it's not going to be reached by more and more billions and trillions of bits of data. That's not the kind of thing it is. We differ from monkeys who have brains very similar to ours, or apes, the great apes. We differ from the great apes only by a few K of code. In that few K of code is the bootstrap program [for] bootstrapping this qualitatively different type of program that we run, infinitely different.

So you asked how optimistic I am. On the one hand, I think that with hindsight, we'll realize that there wasn't much to it. Like, all we have to do is write this program of a few K. And we're done. On the other hand, I see no sign of the philosophy that would allow us to do that. And It's rather like the question of what is life, what that was like in, say, 1800. Some people wanted life to be explicable as an ordinary physical process without any supernatural, without any magic, without any God, just laws of physics. And no one knew how to do that. They had vague ideas, like Lamarck and Darwin's grandfather had ideas that maybe it happened gradually, maybe it happened very slowly. They didn't have the idea of genes, and they didn't have the idea of mutations and natural selection, and that solved it. You could write down that idea in one paragraph.

38:19 **Sean Carroll** It's very easy.

38:23 **David Deutsch** Darwin felt the need to write a whole book, and probably rightly, because from that paragraph, nobody [but] him would have understood it. And it's possible that the idea that will open the door to AGI is that kind of idea. There will come a time when everybody thinks it's obvious and that we in our time were being obtuse for not seeing it. But from this end, it might be very, very difficult.

38:53 **Sean Carroll** But it sounds like it also could be an example of what we started by talking about—we're percolating along in a kind of steady state for a while and then there'll be a sudden change.

39:01 **David Deutsch** Yes. That certainly was the case with Darwin, and it also was the case with Turing. Babbage and Lovelace had the idea, they very nearly had the idea, but they were unable to persuade anybody. They thought it was really important. No one else did. And then Turing, I don't know how long Turing's idea would have percolated if it hadn't been for the Second World War. Although I don't think it would have been centuries, but it might have been [a] couple of decades more before anyone thought of actually making these things. People thought of this as being a bit of mathematics. It's very hard to get into that mindset because we've got computers all around us. I'm wearing one on my wrist. That would have been an alien conception 100 years ago.

40:03 **Sean Carroll** So there does seem to be some similarity here, but you'll tell me whether it's a real one or not, between this idea of our ability to do Turing-complete calculations, explanatory universality. Now, we puny humans can change the universe in a profound way. Does that have anything to do with constructor theory, which is another thing that you have introduced to the world?

40:30 **David Deutsch** Yes. I can't yet give chapter and verse, but I think it's very much to do with it. For example, in the theory of computation, the first thing we work out, or that Turing worked out in theory of computation, is that there's a distinction between functions from the integers to the integers that can be computed and those that can't be computed. Similarly, in physics, we have physical transformations that can be brought about and ones that can't be brought about. So going faster than light can't be brought about. Going to the Moon can be brought about. So there's this distinction. The basic idea of constructor theory is that all the laws of physics can be expressed in such terms, in terms of a distinction between what can be brought about and what can't be brought about. We haven't done that yet. We've basically done it for quantum theory, which was the easiest case. We were kind of modeling the constructor theory on the existing quantum theory, very conducive to it. And my colleague Chiara Marletto has also done it for thermodynamics.

So once we have done that, as it were, or at least conceptually, once we have understood what expressing all the laws of physics in constructor theoretic terms would look like, then the next question to ask, actually I'm already asking it but I'm jumping the gun: Is there a universal constructor? Now, von Neumann asked that question, but that's only because he gave up on the idea. He wanted to have a theory of constructors, what we would now call constructors, in order to understand what life is. This was before DNA and before DNA was discovered and invented, [when] the theory was discovered. But he was unable to. And so he invented the theory of cellular automata instead, and he invented the theory of universal constructors within the theory of cellular automata. But that's not what we want in physics. What we're trying to do is to set up a theory of constructors and of the universal constructor within physics. Then…we don't have a proof, but again, it's very connected with the principle of optimism: Is there a principle that says the things that can be… transformations that can be brought about are precisely the ones that a universal constructor could bring about? And that's, as you see, I mean, you're nodding, so I see you're sympathetic to the idea…this is close to [the] philosophical idea of optimism and so on.

And also, that means that human bodies are a kind of hybrid thing. [Our] brain is both the controller of a universal constructor, which is the human body, because the human body can, or at least in cooperation with others, can build a computer, which can build a universal constructor, and so on. But it's also the programmer. It's also the entity that creatively invents the programs…a universal constructor is not allowed to be creative. It has to be perfectly obedient. So, obedient is the opposite of creative. So the universal constructor is like a universal computer: if it's not going to obey its program, it's not a universal computer. Same with the universal constructor. But our body, as you said earlier, you know, it's imperfect, obviously, It doesn't always obey what we tell it to do. But those are errors which can be corrected. And in principle, these corrections can be achieved with sufficient knowledge. So it's all down to knowledge. So constructor theory is all down to knowledge, ultimately. And same with epistemology and same with everything.

44:52 **Sean Carroll** And we are going to get there, but I guess I would just like to clear up…I did have Chiara Marletto on the podcast before, we had a wonderful conversation. But even though I understood much more after talking to her about constructor theory than I did before, I still think there's this lingering sort of naive physicist's question, which is: If I have a planet orbiting a sun and I know its position and velocity, Newton's laws tell me how to calculate Kepler's laws that [it] will go in an ellipse and things like that. How, or why, should I think about that kind of problem in terms of what can possibly be done and what cannot possibly be done? Why is that a useful or allowed reformulation?

45:43 **David Deutsch** It's not a reformulation. So that type of question, like: What will the planet do? Will it move in an ellipse, that kind of thing. That set of those questions is a subset of those that we really want to know. So for example, we want to know: Are we safe from, to take a thing we mentioned earlier, are we safe from asteroids? Well, for that we want to know: What kind of asteroids can be deflected? Now, existing ways of formulating physics can answer questions like: What kinds of asteroids can be deflected with chemical rockets and telescopes that see the asteroid from such and such a distance and so on. But we're not really interested in that. In the immediate sense we are, but what we really want is to be safe. We want to be able to say protecting the Earth is possible. Then we can work out what kinds of things would be needed, and then we can use the existing-type of laws of physics to work out numerically what will be done here.

But this is different from, say: Can we visit other stars? Well, there we've got a hard limit of the speed of light. So then, if you ask a constructor theoretic question about that, you will immediately come to, "What do you mean by 'visit'?" Some types of visit are possible, some types of visit are impossible, and that is compulsory, provided that the laws of physics are what we think they are. In other words, provided that the dichotomy that the existing laws make between the possible and the impossible is what we think it is. It might not be. Constructor theory won't claim to be the final truth about everything or even anything.

48:06 **Sean Carroll** But I guess the thing I'm still not clear on, then, is…constructor theory might say that a planet can move in an ellipse. Is it supposed to also be a way of figuring out that planets move in ellipses, or does it just say refer to Newton's laws for that?

48:25 **David Deutsch** So it's not constructor theory itself. So constructor theory is [a] kind of meta-law, like the conservation of energy or something. To derive an actual experimental conclusion from the principle of conservation of energy, you have to know what the energy of a particular type of object is as a function of its parameters, it's half MV squared or something, there's kinetic energy. Now, if you didn't know it was half MV squared, the principle of conservation of energy would tell you nothing about how it moves.

48:56 **Sean Carroll** Fair enough.

48:58 **David Deutsch** So that principle is a framework within which theories can be formulated. So if we formulate a theory that violates the principle of conservation of energy, we know that we're postulating something very significant, because we consider that principle to be an overarching principle that governs other laws. Now, constructor theory is intended to be such an overarching principle. So things can be expressed in constructor theoretic terms. In other words, in terms that will say, for example, what can be done to a planet to make it do a certain thing. What transformation can be done to it and what can't. Now, a special case of that is: supposing you don't touch it. What can be done to it without doing anything to it? Okay, but that's a tiny minority of the possible interesting questions.

50:00 **Sean Carroll** One of the, I would imagine, hoped advantages of constructor theory is that it kind of crosses levels, right? I mean, we can talk about biology and chemistry and physics all under the same umbrella.

50:12 **David Deutsch** Yes, very much so. Probably Chiara Marletto already told you that thermodynamics is a prime case of this because in thermodynamics, we really don't want to know what the specific physics of the stuff we're dealing with that do work and [that] have heat and so on. We want principles that transcend that and talk in terms of those. So we want to say for all theories that govern a thing, you can't convert all its heat into work. If you had a theory that violated that, you'd be proposing a momentous thing. You'd be proposing that the second law is false [or] that kind of thing. We're hoping that the same thing will be true of constructor theory, that there will be momentous principles of constructor theory which, on the one hand, will constrain other theories and, on the other hand, will give a deeper understanding of why subsidiary theories—other theories—have the properties they do. We know why kinetic energy in the Newtonian approximation is half MV squared and not half MV cubed. And we know that because we've got a deeper formalism now underlying that which Newton laid down, which Newton didn't know anything about energy. But we know now.

Actually, I think he did. I read somewhere that…you can ask Julian Barbour about this. I think Newton did know a lot about what we now call modern theory of dynamics or Lagrangian or Hamiltonian dynamics, but he decided not to include it because it was irrelevant to what he wanted to show. He wanted to have three laws of motion, better than five. Exactly how you work this out, well, he didn't know the immense power of modern ways of expressing his theory. So I think he wrote down some of these laws. Like Galilean invariance, for example, he definitely knew about.

52:36 **Sean Carroll** Right, that he definitely knew about. It is interesting. So I perceive dimly through the mists a connection, or at least an intellectual affinity, between the idea of separating out possible transformations from impossible ones, and a kind of Popperian epistemology about possible worlds that are allowed by the data and possible worlds that are not. Am I making that up, or is that there in your head, too?

53:03 **David Deutsch** That's definitely there. So, Popper taught us that the content of a scientific theory is in what it rules out. And if you just took that seriously as the basis of your worldview, you'd immediately come to constructor theory because then you would say: What does it rule out and what doesn't it rule out? And the distinction between those two is the theory. Popper never said that. If he had said that, then he could have discovered constructor theory. But yes, it's very much connected, and it's also connected with optimism. Popper's philosophy, he explicitly said, "It's our duty to be optimistic rather than…" I've forgotten the quotation, but it's something like, "Rather than complain about how things are, it's our duty to make things how they ought to be," something like that. So all these things are connected, yes.

54:15 **Sean Carroll** Let me confess, maybe you already know this, but I have long been an evangelist for Bayesian reasoning and epistemology. And I'm fascinated by the fact that you more or less thoroughgoingly reject it, or at least, you know, in certain cases. So explain what your objections are to that because it's subtle but potentially super important.

54:42 **David Deutsch** Yeah, I think it is super important. So quite a lot of different things are called Bayesianism. I don't know which of them you are actually attached to and which kind of come along for the ride. So I specifically object to Bayesian epistemology, which is the theory of knowledge that knowledge consists of propositions in a rational mind, each of which is accompanied by a number, not literally, but implicitly.

55:21 **Sean Carroll** In principle, yeah.

55:24 **David Deutsch** And that these numbers obey the probability calculus. When we say that we've objectively improved our theory, we mean that we've increased the credence, the probability of true theories, and decreased the probability of false theories. So I'd rather not call that thing, those numbers that are supposed to be in the brain, I'd rather not call them probabilities at all. So I try to only call them credences. Because, first of all, I don't think they exist. And secondly, if they do exist, Popper and Miller proved that they don't obey the probability calculus and couldn't. And the key to understanding what Popper and Miller did…as you know, I'm writing a paper about this with my other colleague, Matjaz Leonardis. We have been writing it for years. So It's quite a thing to get one's head around, but the thing we think is the key nowadays is that increasing your credence...

Okay, let me backtrack a bit. If we were talking about logic, then it would be the case that if you prove a theory logically, you've also proved all its consequences. No matter how arcane the consequence, you can't both assert a hypothesis and deny any of its implications. Now, the thing to concentrate on in why Bayesian epistemology is a bad idea is that this isn't true of probabilistic reasoning. So you can have some evidence that increases your credence for a theory. Oh, and now I have to stress that I'm now talking in terms of credences, which I don't think exist.

57:50 **Sean Carroll** We'll let you do it.

57:50 **David Deutsch** Sorry?

57:53 **Sean Carroll** We'll let you do it. We understand the conditional nature of your statement.

57:57 **David Deutsch** Yeah. So I'll say at the beginning that in arguing about Bayesian epistemology, almost every sentence would have to be prefixed with, "Assuming that credences exist and obey the probability calculus, then so-and-so." So the key is that a piece of evidence can increase the credence of a general theory while decreasing the credences of its consequences. And then one has to ask which consequences, because we're only interested in some of the consequences of a theory. It might be that it only ever decreases the credence of uninteresting consequences. And when Popper and Miller proved their theorem, some people took that tack in criticizing it and saying, "Well, yes, it decreases the credence of some of its consequences, but those aren't interesting consequences." But Popper and Miller also proved a criterion for which consequences have their credences increased and which have their credences decreased.

And the answer is: the ones that have the credences increased the most are the ones that just restate the evidence. In other words, the ones whose credences are increased but not that much are ones that are very close to the evidence. And then there are most of the consequences, the ones that are not implied in whole or in part by the evidence. I should say that 'implying in part' is a can of worms because all theories imply tautologies. Therefore a theory and its negation and everything. So there's no way of ripping apart one set of consequences from another.

1:00:24 **Sean Carroll** This is why it takes a long time to write the paper. I get it. Yeah.

1:00:28 **David Deutsch** Yeah. They just wrote down that, they were satisfied with writing down the truth. It only took 3 or 4 pages. They sent it off to I think British Journal for the Philosophy of Science and also to Nature. The papers were accepted. Some people got very angry, and most people didn't notice. And we think that everybody should notice and nobody should get very angry. So their theorem shows that the only way that interesting consequences of a theory have their credence increased is if they have a lot in common logically with the evidence. They're just either restating the evidence or almost restating the evidence.

1:01:21 **Sean Carroll** Is that necessary, or does that happen depending on what your other possible propositions are?

1:01:33 **David Deutsch** The way we prove it is [that] we take the set of all possible propositions expressed in terms of possible universes. [I] quite like that framing. So a proposition or a theory, again, it's a bit like constructor theory. A proposition sets up a dichotomy between the universes whose existence is denied by that proposition—in other words, the universes which couldn't exist if the proposition is true, or couldn't be the real one if the proposition is true—and those that could still be the real one if the proposition is true. You can express it in terms of the set of all propositions or the set of all dichotomies between universes that can and can't exist according to a particular proposition. So if you're a Bayesian—there's my prefix again—if you're a Bayesian, you will want to have a probability distribution function over that set of propositions.

Sean Carroll: Yes.

David Deutsch: And you'll want it to obey the probability calculus. Now, [the] Popper-Miller theorem is independent of what that distribution is, so long as it obeys the probability calculus.

1:02:53 **Sean Carroll** And obeying the probability calculus just means there are numbers between zero and one that add to one.

1:02:58 **David Deutsch** Yes, but there's also relative probabilities. So yes.

1:03:04 **Sean Carroll** That set of ideas. Yeah, Okay.

1:03:06 **David Deutsch** Yeah. The result about the only things whose credence goes up are the ones that are basically restating evidence or something like that is independent of the priors. It's independent of the prior probability, a credence distribution function. Their theorem is true regardless of credence distribution function. Not that there aren't other things very wrong with the idea of a credence distribution function, but at the moment we're assuming Bayesian epistemology. And I should say that other parts of what's sometimes called Bayesianism, for example, the fact that it's a common mistake to use absolute probabilities when one should be using relative probabilities, [and] that's a common mistake that one should avoid making. That's untouched by the Popper-Miller theorem. That's true and, you know, we have no quarrel with that. It's just Bayesian epistemology, that is, the theory of knowledge that says that we obtain knowledge by increasing our credence for true theories. That's the thing that's false.

1:04:20 **Sean Carroll** So is it possible to articulate what explicitly goes wrong with an idea that I would happily tell people? For example, we have two theories of dark matter. We have their weakly interacting massive particles or their axions. And we have some credences on the one theory is right, the other theory is right. And we go out and do an experiment and we rule out some of parameter space, and now we can use Bayes' theorem to adjust our credences accordingly. Is that okay or is that problematic in your view?

1:04:55 **David Deutsch** The way you said it literally is very problematic. What you're informally referring to happens all the time and is perfectly legitimate. And so let me try and say what the difference is. So the picture of knowledge and the growth of knowledge that we have in Bayesian epistemology is that all these propositions are kind of in the frame. We're trying to rule out some of them, increase our credence for others. In real life, what we're seeking is good explanations, and they are very rare. Not only do they not obey the calculus of probabilities, they don't even obey ordinary logic. They don't model ordinary logic. For example, my favorite example, the negation of an explanation is never an explanation. So if you say gravity is caused by the curvature of spacetime. That's a theory. That's an amazing explanation of why we appear to feel forces and all that.

To say gravity is not caused by the curvature of spacetime doesn't explain anything. It doesn't even purport to explain anything. It might be part of your psychological journey from Einstein's theory to quantum gravity or something, but it in itself doesn't tell you anything about quantum gravity. I can prove that to you now because we don't have a theory of quantum gravity that works. So I can prove to you now that merely saying Einstein's theory isn't true doesn't tell you anything about quantum gravity. So that means that if even logic doesn't model what we're doing, then certainly [the] probability calculus doesn't.

And then there's this paradox of the intransitivity of support, as the logicians call it. There's a logician called Hempel who many decades ago proved some theorems. And so, can I quickly explain what [this] intransitivity is? Again, it's just to stress that increasing the credence for a theory does not increase the credence of its consequences typically. Only very rarely does it, and those are the uninteresting cases. Let me borrow the Linda example, [from] Kahneman and so on. You have Linda, who we're wondering whether she's a banker and a feminist. So Linda is going to turn out to be a banker and a feminist. And so, by the way, this isn't the Kahneman thing. I'm just stealing the example. We are interested in the theory that Linda is a banker and a feminist. That's going to be our theory that we're wondering about. So then we find that she's a banker. We get evidence of that. We see her going to a bank every day to work and so on. It increases our credence to nearly one that she's a banker. That will support our theory that she's a banker and a feminist. And once we believe that she's a banker and a feminist, we can go on to deduce logically that she's a feminist. So we've gone by probabilistically from her being a banker to her being a banker and a feminist. And from that, we've gone logically to being a feminist, but her being a banker is no kind of support for her being a feminist. So there's a nontransitivity there. In fact, her being a banker is probabilistic evidence against her being a feminist. That is assuming that the prejudices embedded in that example are true.

1:09:17 **Sean Carroll** For purposes of the story, yeah, we'll go with the prejudices.

1:09:20 **David Deutsch** So being a banker supports being a banker and a feminist. Being a banker and a feminist supports because it implies being a feminist. So A supports B, B supports C, but A countersupports C. And so what Popper and Miller perhaps should have asked at the beginning of [the] paper is: Which implications of a theory are supported by evidence that supports the theory? And they should have said then, "We shall prove actually very few of them."

1:10:06 **Sean Carroll** So I guess I don't quite see the force of this example in this case because I completely agree that the evidence that Linda is a banker increases credence that she's a banker and a feminist, it also increases our credence that she's a banker and not a feminist.

1:10:24 **David Deutsch** Yes, absolutely true.

1:10:27 **Sean Carroll** So I don't see how I would overall increase my credence that she's a feminist just from that evidence?

1:10:34 **David Deutsch** Well, your credence that she's a feminist has increased from what it was before. Now, it's true that your credence that she isn't a feminist has also increased.

1:10:47 **Sean Carroll** That sounds like just a mistake, is that?

1:10:50 **David Deutsch** Sorry, I misspoke. One or other of them will increase, which highlights the issue. If we trust [a] theory more, when and why should we trust its implications more? Note that what we're really after is explanatory theory. That's why the original Kahneman sort of example doesn't work properly because they lure us into trying to think of an explanation by telling us all sorts of explanatory content that is relevant to whether she is a feminist or a banker or both, and then they completely discard it and asked the question, "Is it more likely that she's a banker or that she's a banker and a feminist?" None of the story that we're told before that is relevant to that question.

1:11:38 **Sean Carroll** They were psychologists, not logicians, right?

1:11:42 **David Deutsch** Yes, yes. Well, no, no. We're scientists. Like, we want an explanatory theory. Perhaps in an ideal universe, we would like to have a way of deducing the true theory, but there is no such thing. The only thing we can do is go for explanatory power and go for good explanations. In the case that you talked about, the dark matter and so on, we don't have infinitely many theories. We have a handful of good explanations which are good only insofar as the other theories exist. If the other theories didn't exist, any one of those would be our explanation, would be our sole explanation. And we would go around behaving as if we knew it was true. That's the only kind of knowledge available to finite beings.

1:12:49 **Sean Carroll** But in this case where we have two plausible, pretty good explanations of the same set of phenomena and we have to make decisions about where to spend money testing them and who to hire in our physics departments, how can we not say that we have credences on these different proposed explanations?

1:13:09 **David Deutsch** Well, we can have credences as long as they don't obey the probability calculus. Let me first say, maybe this is relevant. You can tell me whether it is. The Bayesian framework for credences does not allow you not to know something. We don't know which of those theories is true. And we don't expect to get to the final truth even once we do know more than we know now. So we're after good explanations. That means that things we do not know, like are we in an alien simulation? We don't know that. It's meaningless to say that we're gonna give that a credence of 0.5 or a credence of 0.99. Nor, by the way, is it meaningful to ask, "Do we have a credence for Bayesian epistemology? What is our credence for Bayesian epistemology? Is it 1 or is it 0.5 or is it 0.99?" Now I can remove the prefix and I can say none of those things make sense. We decide between theories of dark matter or theories of epistemology according to how well they explain what we want to explain. So when we're asking which one we want to fund, which theory we want to test next, we're asking not our credence for the theories, we're asking for judgments about what the prospects for increasing knowledge are.

So I think that quantum theory is definitely false. I think that general relativity is definitely false, and I also think they contradict each other. And therefore, my credences, like if I talk about my beliefs for those theories, they definitely don't obey the probability calculus. Because if they did, my credence for one would be one minus my credence for the other. And yet I have a very high credence for both of them. So probability doesn't provide a proper model for my attitude towards theories. And it's the same with the different theories of dark matter. What we want to do is to do an…ideally, we'd like an experiment that is a crucial test between two of them, after which one of them would have zero credence. So it's not a matter of credences going up and down. Really, credence, provided you are confident that the experimental setup is right, Duhem and Quine pointed out that we can't always be sure of that, and in fact, ultimately, we can never be sure of that because [there are] always the aliens with their virtual reality machine that might be misleading us.

So probability doesn't come into any of this. We want to take into account things like, "How good an explanation was it in the first place?" Like, "If it's a good explanation, can we rule out a bad explanation so that we don't have to consider it anymore?" Or, "Can we fail to rule it out?" in which case we'll have to consider it more than before. "How expensive are these experiments?" We cannot work out how much money to spend on testing each [experiment] by using classical decision theory and seeing which one has the highest expectation value of the benefit that we will get from knowing things or not knowing things because we don't know what the outcome is going to be. The best thing that can happen in an experiment is that you get an outcome that you didn't foresee. But if you didn't foresee it, you also didn't foresee its probability. What is the probability that doing [an] experiment on knocking a comet out of the way will tell us something about dark matter? Well, we don't know, but we don't know that probability, and that probability is irrelevant. What we can use is our best explanation. We can see that none of our explanations of dark matter say that it will affect comets. And if one of them did, then we would ask, "Well, can we test this?" Or is it the kind of theory like, "Well, it could be so," which is always true but is a bad explanation.

We judge on the explanation, not the probability. So if you're alive at the time of Kepler and Galileo and those people, then you're not looking for a high-probability theory. The theory that the planets move on epicycles has got a far higher probability than that they move on ellipses because an ellipse is a kind of epicycle. So by the Linda argument, Galileo should have preferred the epicycle theory because it's far, far more probable than the ellipse theory, but he didn't. He preferred the circle theory, which is even less probable than the ellipse theory because, given what he thought he knew, it was a better explanation because if it's an ellipse, then you've got to explain more things. There's more things [that] remain unexplained than if it's a circle. So ellipse, what's the eccentricity of the ellipse? With a circle, you don't have that question. So he thought there's gonna be a way of making circles work. See, he wasn't looking for [a] high probability. If anything, he was looking for the lowest possible probability that's still viable as a theory.

And that's what we do in science when we're looking for general theories. It's a bit different when we're looking for a particular theory. Now, this comes back to other uses of Bayes' theorem. If you're a doctor and you want to know whether a particular patient has got dengue fever or something, and you ask them, "Well, have you been to the Far East lately?" Then you're asking for something probabilistic. If I bend over backwards, I can call that probabilistic. It's really that he's looking at frequencies, first of all, not probabilities. He's looking at: there's only a finite number of people that have been to the Far East, a finite number who have got infected with dengue fever. He's approximating those frequencies and probabilities, and he's using the approximation that his putative patient is randomly chosen from the set of all those people, which he wasn't. He wasn't, but he's using that because he doesn't know. But does that mean that he's giving [the] "doesn't know" credence of one half? No, he's using it because he doesn't have an explanation of the patient's contact with dengue fever that doesn't include going to the Far East. Now, if they said, "Well, I haven't been to the Far East, but I have been to a lecture that was attended by scientists who've recently been to the Far East, then that would change the priors." Okay. We call those the priors, but [it's] actually just changing the numbers in these frequencies. So it's sometimes a good approximation to approximate frequencies by probabilities, or rather by numbers that obey the probability calculus. They don't increase our knowledge. You can't increase general knowledge that way. You can't decide between general theories in that way because the set of individuals is infinite there. So it won't work there.

1:22:08 **Sean Carroll** It's clear that the idea of a good explanation is kind of crucial here. How clear and formal can we be about what is a good explanation?

1:22:19 **David Deutsch** Well, as we agreed earlier, you can't formalize the concept of a good explanation.

1:22:28 **Sean Carroll** You know it when you see it?

1:22:31 **David Deutsch** No. So it's not like a matter of taste. It's a matter of philosophy. So we can make progress in philosophy by the same method, that is, by saying that we're going to exclude solipsism because solipsism could explain anything, could quote "explain anything." And we're going to exclude the doctor saying, "Well, the patient could be lying, could have been anywhere, therefore

I don't know and I've got no way of assessing whether they've got dengue fever or not." You also exclude that because that is always true and would always short-circuit any kind of trying to approach the truth, but trying to approach the truth about general theories means that you have to adopt the criterion of a good explanation because, well, this argument that an explanation that can explain anything is a bad explanation, I think has got a transcendent compellingness about it. Which doesn't involve any axioms. Like, we're not making an axiom of using the best explanation because if you make an axiom, you'd want to have a precise definition of the terms in the axiom. But somebody, like I said earlier about principle of conservation of energy and that kind of thing, if you want to say that bad explanations are actually acceptable, you've got to realize that you're climbing up a philosophical mountain by saying that. You can't just say that just to justify your own theory to say that actually mountains don't exist because anyone could say that about anything, and if you say, "Well no, although anyone could say it, I'm saying it and there it's allowed." Well, that got an obvious flaw in it, that way of arguing.

1:24:53 **Sean Carroll** That was very helpful, but one thing you said along the way, I can't quite let you get away with, or at least I want to hear more, namely that you're pretty sure quantum theory is false.

1:25:05 **David Deutsch** Yes.

1:25:06 **Sean Carroll** In what sense do you feel that?

1:25:10 **David Deutsch** So, pretty sure, and I'm not saying I've proved it. Several things. The main one is what I mentioned about [its] conflict with general relativity. In general relativity, we know that the behavior of an object, like a planet or whatever, is dependent on the behavior of another object like the Sun, and that this is mediated by a field which travels at finite speed. And quantum theory tells us, with equal confidence, that the Sun isn't just in one place. The Sun is in a superposition or more generally in a mixed state, where it has many different positions simultaneously. And although some of them are pretty close to where we see the Sun, some of them are a long way away. We know that because of the instability of classical mechanics that the Sun has been involved in lots of interactions and some of them will have been chaotic, and therefore will have the end result [that] will have depended sensitively on the initial result. Therefore, these positions of the Sun that were initially very close to each other will get very far away. And therefore, according to quantum mechanics, some of the Suns are far away and relativity does not, and neither of them have a way of telling us that the Sun's effect on planets is different in different universes. I can say that in words, but I can't say it in equations. And therefore, we don't have the right equations.

1:27:08 **Sean Carroll** Would you not? I get everything that you said, but then I want to just say there are different branches of the wavefunction where there's a good semiclassical approximation and general relativity works pretty well.

1:27:20 **David Deutsch** So when I say that the theory is false, I mean it's not true. I mean, I'm using the words 'true' and 'false' in the sense in which they're used in logic. There is no excluded middle. Certainly both relativity and quantum theory are extremely good approximations in the situation where we want to apply them. It's not so clear that we won't very soon be applying them in other situations like in the early universe, where we want to explain something like the distribution of microwave background radiation over the sky, where there are billions of light-years involved. And this is all due to something that happened on a scale smaller than an atomic nucleus originally, where definitely quantum effects were dominant. And we don't know what those were and how they affected gravity and dark matter and spacetime and so on. How close a theory is, how good an approximation is, depends on how you want to use it. How good an approximation a theory is. So yes, certainly good approximations for practical purposes, but so is Newton's theory. That's also false.

1:28:54 **Sean Carroll** Do you have any hints as to how to modify quantum theory to make it better?

1:29:01 **David Deutsch** Yes, I think so. There, I would have to go to quantum field theory, which has more of [an] internal problem, never mind gravity, just the problem of quantum field theory. All existing quantum field theories are based on axioms, which include the axiom that fields that are spacelike separated commute with each other. Now, that also means that a field at one point commutes with the future light cone of the field at the other point. But on the other hand, field quantities at the same point fail to commute. Therefore, field quantities are discontinuous everywhere. So the whole conceptual framework of quantum field theory is not what it's cracked up to be. Now, mathematicians say, "Okay, well, it's not a real-valued field, it's not a quaternion valued field, it's a…" I've forgotten what they call it. But anyway, the only things that are real are the integrals of the field over finite size.

1:30:33 **Sean Carroll** Right. Distribution[-valued] fields.

1:30:35 **David Deutsch** That's what they're called, yes. Distribution-valued fields. But that hasn't got a conceptual model. I mean, you can't have a distribution over things that don't exist. So you can say that only the distributions exist, but a distribution has to be over something. And so anyway, in short, I have an idea for a variant of quantum field theory where we don't have that axiom, where fields at spacelike separated points are allowed to fail to commute, and where the thing that they have to do is be continuous. And [there's] quite a nice theory. Again, mathematically, it's quite nice, but conceptually, it's wild. I rather like it..

Sean Carroll For that reason.

David Deutsch: Yes, yes. So not only do causality and that kind of thing mean a different thing in that theory, but measurement does as well. And the separation of systems into subsystems means something else than it does in ordinary quantum theory. And so I've been trying to get that theory to work for years. And I got some nice equations of motion for it, which I don't know what they mean, but it's rather a nice thing. Because of this pathology in quantum field theory, it's been taken for granted that the way you judge proposed quantum field theories is by how well they let you get [around] those pathologies, whether you have an infinity that cancels another infinity, and so "This discontinuity is not as bad as you might think," and so on. And on the other hand, the ones that don't have that property are not really considered.

Now, in this unorthodox quantum field theory, as we call it, you have a different criterion, and the criterion is simply that the algebra of the quantum fields does not change with space and time. Which we have in the conventional theory as well, except that that hardly makes sense when it's discontinuous everywhere. And then you see that there are only a finite number of possible second-order equations of motion. And so that can be the criterion of the ones that are useful to investigate physically. And I and another colleague, Sam Kuypers, have been investigating [an] easier version of that, where it's just the qubits that don't have to commute. Different qubits rather than field, which is an unwieldy thing. So you have qubits which don't have to commute with each other. And that is another rather nice theory, and it's promising in various ways and we are working on whether this could be testable. Like if you have, say, a pair of photons or something coming off a decay process, whether those two photons might not commute with each other. And if they didn't, could we detect this? It would produce a kind of entanglement between them that is different from the entanglement that happens in ordinary quantum field theory. So we haven't got there yet, but that's the kind of fun we've been having.

1:34:37 **Sean Carroll** Yeah, that does actually sound like fun. Does it fit into an Everettian kind of formulation of quantum theory?

1:34:46 **David Deutsch** Yes, Everettian and in the Heisenberg picture. I and we think that the Schrödinger picture is very misleading because the Schrödinger state is global and it leads Everett and DeWitt to thinking about the whole universe as splitting every time the state changes.

1:35:09 **Sean Carroll** Yeah, I'm all in favor of that.

1:35:14 **David Deutsch** I think that's too much for many people to swallow, and they don't have to. Because in the Heisenberg picture it's only the observables [that] fit in and the distant universe is left unchanged by quantum phenomena.

1:35:28 **Sean Carroll** Maybe this gets into something I've always wanted to ask you about. I think of worlds in the Everettian quantum theory as arising from decoherence, but I've heard you say things that make me think that you're more willing to talk about multiple worlds even before decoherence has happened.

1:35:50 **David Deutsch** Yes. In my view, because I prefer to think in the Heisenberg picture where everything is local. So there are two situations where it is a good approximation to think of quantum systems as splitting into worlds. One of them is when there's decoherence, but the other one is where there is a quantum computation in progress, but not just any quantum computation. If you have a typical quantum computation, in fact, you have a set of worlds that are all identical, then you do something to them that makes them different, like makes them two to the N of them, and the register holds a different number in each of the two to the N universes. Then you do stuff to those numbers in those registers and then you recombine them. So I think that during the process of splitting into multiple copies and in the process of recombining, it's not useful to think of them as being separate universes. They're all affecting each other so much that a universe conceptually is a quasi-autonomous thing. It's a thing where classical laws almost hold. And that's what happens during this intermediate thing where you are doing a different computation in each of a vast number of universes. The computations are classical computations. And they're not affecting each other. Each one is autonomous. There, it's useful to speak of the multiverse as having split into universes for a while. And also when there's been decoherence, it's also useful for, like, the opposite reason because there's no hope of recombining them.

1:37:55 **Sean Carroll** Well, I guess yeah. This is very helpful to me because I get it now. So in the quantum computation case you say it's useful to think about them as separate worlds because they're evolving independently, even though there's probably some sense in which they're not classical, I think.

1:38:16 **David Deutsch** They're classical computations.

1:38:18 **Sean Carroll** But they are classical computations. And furthermore, they do recombine at the end of the day, unlike the decoherence example.

1:38:24 **David Deutsch** Well, if somebody knocks over the computer and they never recombine, and then that happens later, then you can't say, "Well, retrospectively, they weren't universes." I think that wouldn't make sense.

1:38:37 **Sean Carroll** Okay. All right. Well, you've given us a lot to think about. My last question will be: Am I right that you recently mentioned that you're working on a third book?

1:38:47 **David Deutsch** Yes, actually. I'm working on several books, and I'm not sure what I can say about the ETA of any of them. So I'm also working with Sam Kuypers and Chiara Marletto on a textbook of quantum mechanics, quantum theory, and I'm also working on a science fiction book, which contains conjectures that I wouldn't dare state seriously even in an article. But in science fiction you're allowed to.

1:39:22 **Sean Carroll** But maybe you have a little bit of sympathy for these conjectures?

1:39:26 **David Deutsch** Yes, I have a bit of sympathy for all but one, which is very horrible. And that's what makes it dramatic. I don't know how to refute it. I mean, it could be true. But as we've just said, lots of things could be true.

1:39:46 **Sean Carroll** Lots of things could be true. And for the quantum theory textbook, is that supposed to be a competitor for a standard second-year in university?

1:39:54 **David Deutsch** It's a competitor in the sense that, if somebody wants to change the entire way they teach quantum mechanics, then this would be a way of doing it. So [the] Heisenberg picture would be central, not Schrodinger. Everett would be central. Qubits would be central, not hydrogen atom. So it's all about quantum information. It's close to modern kinds of experiment instead of old-fashioned kinds of experiment. And conceptually, it doesn't have the baggage that existing things do. Now, I know there are a couple of textbooks already on the market that start with qubits. And I haven't actually read one of them. But I'm sure they don't do those other things that we want to do.

1:40:57 **Sean Carroll** Probably not. I will confess I'm also working on one very slowly. But I don't know how to characterize it. I'm not as ambitious as you are about hoping that people will completely change how they teach quantum mechanics. Even though I think that I'm sympathetic to the philosophy you put forward in this book, I'm guessing, but I'm going to try to split the difference so that more old-fashioned people are not quite as shocked. So a little bit of everything there.

1:41:31 **David Deutsch** Yes. Well, that'll probably sell much better than our one.

1:41:35 **Sean Carroll** No, I'm not averse to that. We'll have to see. But David Deutsch, thanks so much for being on the Mindscape Podcast.

1:41:41 **David Deutsch** Well, thank you for inviting me.

ARJUN KHEMANI:

FREE-WILL, TAKING CHILDREN SERIOUSLY, AND ANARCHO-CAPITALISM

About the interviewer: Arjun is a high school dropout who launched and helped lead support at Naval Ravikant's Airchat. He is creating a documentary about our deepest theory of knowledge. In general, he works to spread optimism and fight the enemies of civilization. He also works to bring Zcash to the world.

Interviewer homepage: https://www.arjunkhemani.com/

YouTube channel: https://www.youtube.com/@arjunkhemani

EPISODE DETAILS

Date January 1, 2024

Interviewer Arjun Khemani

Source YouTube

Episode #25 – David Deutsch: Free-Will, Taking Children Seriously, and Anarcho-Capitalism

Description
0:22 Happiness is a state of continually solving one's problems
4:54 Both free-will and the self exist
12:06 The principle of optimism
17:28 Any ultimate explanation is a bad explanation
20:22 The origins of Taking Children Seriously
25:33 Why children are the most coerced members in society
31:33 Anarcho-capitalism

Link https://www.youtube.com/watch?v=Hh9xOB2oDHk

Ideas

- Nothing can be proven by experience.
- Einstein's theory of relativity was not contained in the Big Bang. It was not contained in the thoughts of physicists at the beginning of the nineteenth century, but it was created in the mind of Einstein. So humans create things, create novelty all the time in their minds. And therefore it's meaningful to ask whether a particular idea was created by them or by somebody else or by nobody, whether just sheer chance.
- Anarcho-capitalism is a state where no monopoly of violence is considered legitimate. But that can only exist when the problems of not having a monopoly of violence have been solved somehow by some institutions. You can't be in favor of just outcomes without specifying the institutions that would cause it.

Topics all knowledge is conjectural • anarcho-capitalism • bad explanations • bucket theory of the mind • defining things • education • epistemology • evil • foundationalism • free will • happiness • incrementalism • institutions • institutions of consent • J.S. Mill • John Locke • knowledge creation • liberalism • means of error correction • morality • Popper • solving problems • suffering • Taking Children Seriously • the self • transmission of memes

Episode title and description reproduced verbatim from the original source.

Transcript

0:00 **Arjun Khemani** I thought we could talk about some of the wider implications of taking our best understanding of epistemology seriously. Since it is universal for all kinds of knowledge and knowledge creation, Popper's epistemology has some profound and far-reaching implications. It is also the only kind of philosophy that I know of that is actually very practical. In Chapter 12 of *The Beginning of Infinity*: A Physicist's History of Bad Philosophy, you write a bit about happiness, and you conjecture an explanation about the cause of human happiness. You say, quote, "Happiness is a state of continually solving one's problems. Unhappiness is caused by being chronically balked in one's attempts to do that. And solving problems itself depends on knowing how. So external factors aside, unhappiness is caused by not knowing how." Can you please expand on that? It seems that this theory is a special case of the principle of optimism, which you define as 'all evils are caused by insufficient knowledge.'

1:03 **David Deutsch** Yes. First of all, I think you have to start, in answering that specific question, with suffering, human suffering, because that's really what all evils are. I mean, some people would disagree. Some people would say that the Earth can suffer. On the other hand, I think it's not the case that reducing or abolishing suffering is the definition of morality. It's just that they're connected via epistemology.

1:39 **Arjun Khemani** Yes. So I just wanted you to expand on your definitional view on happiness, which is a state of continually solving one's problems.

1:51 **David Deutsch** Yes. Well, I'd rather not define things. I'd rather say there is an issue, there is a problem about what we should aim for and what is right and wrong and so on. There's a whole constellation of problems that arise from the fact that we are capable of creativity and of making creative decisions. And one way of putting that is that we seek happiness, but we don't seek happiness in any—not always, anyway—in any straightforward way, because somebody might decide to join in a war because they think it's right. And they know that there is [the] risk that they will suffer as a result of this. Now, you might say, "Yes, but they would suffer even more if they refused to go because they wouldn't be thinking of themselves in the same way and as the same kind of person," and so on. Yes, but that is approaching the problem backwards. There is a reason why they want to think of themselves in one way rather than another. And that way of thinking about themselves is rooted in their theory of morality. So what they would be happy or unhappy doing depends in part on their morality. It also depends in part on other things, like their environment and their culture and their memes that they've inherited from other people and so on.

And all those things, ranging from what they have thought of themselves to what is hardwired in their genes or whatever, all that is mutable. So they can in principle change all those things. And if they're changing them successfully and aren't thwarted or balked in doing that, then I think we can call that happy to a good approximation, even if they are, in the more superficial sense, suffering as a result. So, yeah, I should have begun by saying, like Popper, [that] I'm not too keen on defining things. The question is always: What problem are we faced with? And as [with] defining things, you know, what concept comes up again and again, when we're thinking about particular kinds of problems, and is it worth giving those a name, and then we can call that happiness. But if you say, "Yes, but it doesn't conform to the usual way of using the word 'happiness,'" I mean I often use the word 'fun.' Well, in that case, I'm happy to use someone else's terminology. So long as their terminology does not define away distinctions that I want to make in discussing problems.

4:54 **Arjun Khemani** Some people think that it is the transcendence of the self that gives rise to true happiness or to enlightenment. They say that the self is an illusion and free will does not exist. What do you think about this claim, which some people think can be proven by experience?

5:13 **David Deutsch** Well, nothing can be proven by experience. So that's [an] easy one. I think those two things are different, the existence of the self and the existence of free will. I think they both exist, but they are, I suppose, connected, but there are differences between them. I think, for example, it's true that when we are deeply engaged in a problem, an absorbing problem, then typically one does not think of oneself while conjecturing and criticizing in regard to that problem. You know, notoriously people forget to eat when they're immersed in the problem. I was just reading a Sherlock Holmes story yesterday, forget which one, in which Holmes remarks that he hasn't eaten since yesterday or something. So that's a notorious thing and I think it's true.

6:09 **Arjun Khemani** It's a sign that you should be doing what you're doing.

6:13 **David Deutsch** Yes, yes. It's a sign. It's not an infallible [sign]. Nothing is. Now, that doesn't mean that the self doesn't exist, though. You know, sometimes we think of ourselves qua self, and sometimes we think of the early universe. And when we're thinking of ourself and not the early universe, that doesn't mean that the early universe didn't exist. It existed, we're just not thinking about it. Similarly, the self exists, and sometimes we don't think about it. It is very hard to define because it is connected with consciousness, which we don't have a very good theory of.

Now, free will—it seems to me that a lot of the discussion about free will is really literally a discussion about nothing because people define free will as a type of thinking that violates the laws of physics. And then they say, "But you can't violate the laws of physics, therefore, there is no such thing as free will." Well, there's no such thing as free will thus defined. Yes. But I think the way free will is used both in everyday life and in philosophy isn't like that. It's got nothing to do with violating laws of physics. We need that concept in issues like: When a person X kills person Y, did they do that of their own free will? Or was it that a gust of wind pushed them towards the other person and then the other person fell into the path of the train, and therefore that wasn't initiated by them. And then you can go into detail, like saying, "Well, if they have a hateful state of mind and that was inculcated in them by their parents, are they really exercising their free will when they commit the crime or putative crime or when they enact their parents' crime?" However you want to phrase it. Well, that's a matter of fact. It's a matter of fact whether the process that led from their parents' inculcation to their pushing the person into the path of the train, whether that ever involved [a] process in their mind that is worth calling free will.

So in that kind of situation, when we say, "What kind of process is worth calling free will?" we really mean, "What kind of process is worth considering to have interpolated between the parents' inculcation and the pushing action in such a way that it shifts the moral responsibility?" And I think the answer to that, or a big part of the answer to that, is whether the person in this thinking, during this thinking, has created something new in the world. New knowledge, but it might be false, but whether they have created something.

Now, again, many people would say there is no such thing as creating something new in the world. Everything you create, not only is it called by your parents inculcation or whatever, but it's caused by the laws of physics. It's caused by the state of the universe as it was a hundred years ago, as it was billions of years ago, as it was at the Big Bang. So the Big Bang is what caused this act of pushing. Now, I think that's an incoherent view of novelty in the universe. It is clearly not true that—I say 'clearly' advisedly—someone's actions today are caused by the [state] of the universe yesterday, because all the arguments that it is apply equally well to tomorrow. So you could say, "He pushed him into the path of the train because he's going to be sitting in a jail cell tomorrow," and that has the same logic. The laws of physics are time-reversible. So this kind of arguing away of the existence of novelty, and therefore of free will, is incoherent. It's literally infinitely ambiguous. And some of the ambiguous ways of deploying that argument are mutually contradictory.

And common sense says that Einstein's theory of relativity was not contained in the Big Bang. It was not contained in the thoughts of physicists at the beginning of the nineteenth century, but it was created in the mind of Einstein. So humans create things, create novelty all the time in their minds. And therefore it's meaningful to ask whether a particular idea was created by them or by somebody else or by nobody, whether just sheer chance. And it can be a matter of degree. You can say the idea was partly created by Einstein and partly by Lorentz and Hilbert and so on. And it might be hard to tease out exactly what idea was created by Einstein, but you can tell it was because those people did not write those papers.

So yes, so I think both free will and the self exist and the arguments against them are no good.

12:06 **Arjun Khemani** In your most recent book, you explain that a momentous dichotomy exists, which is that any physical transformation is either impossible because it is forbidden by the laws of physics, or it is achievable given the right knowledge. And the principle of optimism is that all evils are caused by insufficient knowledge. Doesn't this principle assume that the things that are forbidden by the laws of physics are not evil?

12:34 **David Deutsch** Yes. Logically, they could be. It could be that the universe is evil. Certainly, the universe is not as friendly to thinking beings as a lot of people think it is, but it's not as evil as a lot of other people think it is. I mean, it's not evil in that sense in that, well, I think it's best to think of it as indifferent rather than evil. It doesn't care whether we're happy or miserable. There's nothing in the universe that's directed towards our being happy or miserable. As I say, logically, you could imagine a universe which was like that, where it was directed towards human affairs, just like religions sometimes often say that "There are supernatural things in existence that care about whether humans are happy or not." And sometimes they are cast as wanting humans to be happy and sometimes vice versa and so on. In regard to the modern scientific worldview, this will be very surprising because the facts of the universe are arranged according to laws of physics, which are highly universal. They apply to the planet Venus and to formation of galaxies many billions of years ago and to events in the universe that haven't happened yet and so on. Nothing about [them] is specialized to our planet or to our species. So it would be very odd if somebody found one day that the real laws governing the electron are the Dirac equation plus an extra term that we hadn't noticed that said that if Arjun is happy, thwart him in such-and-such a way. It wouldn't fit in with the way we now know the universe runs. You can't rule it out logically, but there is no motivation for assuming such a thing.

14:53 **Arjun Khemani** As you know, bad outcomes can have good intentions tied to them, or perhaps they can even have no intentions tied to them, right? So I'm trying to think how even if the universe or the proverbial asteroid or something, it is a problem to us, to us humans or to any human in particular, [or] to a person, then is that not an evil because the universe is indifferent to us?

15:22 **David Deutsch** Yes. If an asteroid turns up in such a way, I think by now it would have to be a bit of a minor planet. If it turns up in such a way that we haven't noticed it, and it's going to wipe us out in too short a time for us to prepare for it or to deflect it, then that was caused by a lack of knowledge in the sense that we could have generated that knowledge. And our slowness in generating that knowledge is itself due to a lack of knowledge. We didn't have that knowledge for a thousand years during the Dark Ages. And even before that, we only had it in certain subcultures and so on. We have existed as a species for two or 300,000 years, most of which were wasted from the point of view of creating knowledge. So you know, maybe in some universes we were wiped out, or presumably in some universes we [were] wiped out early on, before we even in principle would have had the chance to create the knowledge.

I think, yes, I think that would better be called 'indifference' than 'evil' because the process that brought that about was an unlikely one. And it wasn't tuned to causing human suffering or death or whatever. And we see that the Earth has been struck, and there's nothing in that event that could be called 'targeting.' That is there's nothing in that asteroid or where it came from right back to the beginning of the universe, right back to the Big Bang, there's nothing in that process that knew about dinosaurs and wanted to wipe them out. So it was an accident.

17:28 **Arjun Khemani** You've said before that if all knowledge is conjectural and subject to improvement, then protecting the means of improving knowledge is more important than any particular piece of knowledge. This goes against so many common assumptions about what knowledge is, what morality is, and how we know things. Some people think that we have to reason up from a few axioms or a foundation that we know to be unquestionably true. But you think that any ultimate explanation or foundation is a bad explanation, why is that?

18:04 **David Deutsch** Yeah, so a bad explanation is one that is easily varied. And if you have an explanation that cannot be questioned, then that means that the question of why the unquestioned thing should be that rather than a different unquestioned thing, which somebody else advocates as being unquestionably true, is unanswerable. And so that's the very epitome of a bad explanation because if somebody believes a particular version of this and a different person believes in a different version of it and comes from [a] different culture, then they have nothing to offer in terms of a reason for why they're adopting their view. And the structure of their explanation of morality, or whatever it is, applies equally well to the other person.

Some people say, "Well, what [if] we have an explanation that doesn't need any assumptions, which just is self-evident, and nobody has a rival theory?" Well, if nobody has a rival theory, then we don't need morality, like everybody will agree. But, first of all, there is no such thing. And secondly, even then, even if everybody agreed, that still wouldn't make it true. It would still be important to criticize that thing and try to discover a deeper meaning to why it was true. Why is it that everybody agrees? Is it perhaps because of a widespread irrationality inculcated in them, or is it because it's really true? And because it's really true, that possibility requires some critical thought about this unquestioned thing, even if true. I think J.S. Mill said something like this. No doubt he said it much better than I could, but something like, "Even if you're absolutely right, unless you," I can't say it properly, but "unless you know why your opponent is wrong, you haven't really understood why you are right." And that requires criticism of both your idea and the opponent's.

20:23 **Arjun Khemani** After reading *The Beginning of Infinity*, I started quoting you everywhere. I was writing a blog post about parenting, and I quoted you about something against authority. And I extended that argument to education and to children. And at that point, I hadn't heard of Taking Children Seriously, but I just naturally seemed to overlap with those ideas. And our mutual friend, Brett Hall, read my piece, and he told me that—in maybe no accident—I was quoting you because you founded a non-paternalistic movement with Sarah Fitz-Claridge called Taking Children Seriously. Then I dug up the content online, and it all immediately just made sense because it was all an application of Popperian epistemology to children and to education and parenting. So I thought we could talk a little bit about Taking Children Seriously, starting with: What was your and Sarah's motive for starting it as a movement?

21:32 **David Deutsch** Well, I can't really speak for Sarah. I think our motives were different. Mine came from Popper, and I noticed… well, I suppose it came in some stages, but they happened quite fast, you know, one after the other. First of all, realizing, as Popper did, but I didn't really get it at first, that his epistemology is very general. He applies it at great length to political philosophy and to science. He didn't ever apply it to economics, for example, and he didn't apply it to education. At least, he hardly ever wrote about education. That was because he didn't want to write about psychology because he didn't want to give ammunition to people who adopt subjectivist theories of truth and knowledge. But occasionally there's a little gem here and there in Popper, where he says the right thing, or more or less the right thing.

So, having realized the generality of this epistemological theory and having been convinced that its opponents are wrong, I said to myself, "Well, if this epistemology I was just reading is true, then everything in existing educational theory is false." It simply destroys it, and it destroys it with [a] sort of philosophical firepower that's overwhelming because the existing education theory is very parochial. It's centered on alleged properties of humans, and it's mixed with theories of psychology and so on. And yes, certainly one does need theories of psychology in order to understand education. But if one violates epistemological principles, then one is wrong. The arguments for why one is wrong are just too powerful. So then I started thinking about it, and that was the beginning of my journey into educational theory. And I found that a lot of this realization was not at all new. There's a long history of what you might call noncoercive educational theory: Rousseau, Godwin, and, in the twentieth century, Montessori.

24:00 **Arjun Khemani** You might have some quibbles with that, right?

24:02 **David Deutsch** Well, I have quibbles with all of them. But I think all of them have some things in common. And what they have in common is that they view educational theory as a branch, a small branch of, they would put it in different ways, but I would say of epistemology or of philosophy or of liberalism. Even John Locke realized this, even though his actual educational theory was awful, but John Locke was a long time ago. And by the time it came to Godwin, it was pretty far advanced. Though even he was awful in some places, and Rousseau, and so on. So I think putting it all in a more…I used the word 'powerful' just now. It's not quite the right word, but anyway, in a proper philosophical framework, changed all that into TCS. And really, it didn't require any more than Popperian epistemology. I don't know what he would have thought of it. He was just this guy, you know. His theories were the best we have so far in epistemology and related things, but he made many mistakes, and maybe he would have made a mistake or two there as well. So anyway, that's how I came to it, via Popper.

25:33 **Arjun Khemani** Is there an explanation for such widespread coercive control over children in society at large?

25:43 **David Deutsch** Yes.

25:43 **Arjun Khemani** Do you think [there's] something that explains that?

25:45 **David Deutsch** I think basically, if you have a conception of knowledge, well, one can put this in several ways. As Popper would put it, if you have the bucket theory of the mind, which is that knowledge is like a fluid that exists in one generation and can be poured into the next generation mechanically, then you will automatically have a coercive theory of education. You might be very kind about it, you might be harsh about it, or you might be anything in between. You might have idiosyncratic quirks. But you will basically be wrong because, you know, that's just as bad, as just as error-prone as trying to—my other metaphor, simile—it's like trying to fly by jumping high. It's just the wrong thing. It's the wrong picture of what the problem is. And the right picture of what the problem is, because of this, the right picture of what the problem is highly counterintuitive.

It's like, whenever I criticize a particular feature of existing educational practice, like, say, exams, then the natural thing is to say, "Well, how will the knowledge get into them? Okay, some of them will want to learn that, but some of them won't." And what about those? And this, "some of them won't and some of them will," that's all the wrong theory. It's all the wrong picture. Like, if somebody doesn't want to learn something on the curriculum, it's because they are learning something else. They might be making a mistake, or you might be. And the way to solve that kind of issue is to discuss it and to bring criticism to bear and to respect, not only the principles of epistemology broadly, but specifically the principles of liberalism, which already tell us…we're lucky, but for hundreds of years now, we have had liberalism, which has already solved the apparently intractable problem of how you get people who have different views about things to live together without violence. Now, that's not enough for an educational theory, but it's a thing that an educational theory ought to respect and conform to. And if we still have the educational theory that we had in pre-liberalism days, then that's a sign of something that needs correcting just in itself. It might be that certain practices are justifiable under liberalism as well. It's an issue that needs to be addressed.

The reason, now this is a speculation, I think what I just said is fairly straightforward, but the reason why it's especially intractable in our society is that it is connected with the transmission of memes. So education as conceived at present is entirely the transmission of memes. Now, without transmission of memes, we would be reinventing the wheel all the time. So it's highly desirable that memes be transmitted. It's also highly desirable that they not be transmitted faithfully, because otherwise there won't be improvement. And if there's anything worse than having to reinvent the wheel, it's having no improvement because, whereas reinventing the wheel all the time is terribly inefficient, not having improvement is fatal. It's certainly fatal. So we need to solve this problem. And liberalism has solved it in the political sphere. Capitalism goes a long way to solving it in the economic sphere. The canons of rationality in science, especially as improved by Popper, have more or less solved this in the scientific sphere. But in education, because educational theory is [conjectured] in the context of the wrong problem situation, it's especially resistant. This is all speculation.

30:22 **Arjun Khemani** Right. I think with parenting and with most other things that are worthwhile, it just requires creativity. And I see a lot of people with [the] zero-sum mindset that if the child gets what they want, then the parents can't get what they want. Or there's a compromise and nobody gets what they wanted. But there can be a solution and it just requires creativity.

30:50 **David Deutsch** Yes. And it is interesting that in regard to liberalism and capitalism and so on, a few hundred years ago the analogous arguments were being made to say that liberalism is impossible, democracy is impossible, because, you know, "What might the people vote for? What if they voted for all institutions to be overturned? What if the majority voted for the minority to be dispossessed?" Well, that can happen, but under the institutions that have evolved, that doesn't happen. That's not the problem. Democracy has plenty of problems, but that's not one of them.

31:33 **Arjun Khemani** This ties to the next topic I want to talk to you about, which is anarcho-capitalism. From my understanding of the ideas of Popper and yourself, a system of anarcho-capitalism really does seem like an ideal system for maximizing knowledge growth and progress in society. What are your thoughts on anarcho-capitalism? Do you think it is desirable, or do we need certain institutions like the state to maintain stability? Feel free to talk as much as you like about it.

32:08 **David Deutsch** Well, I don't think anarcho-capitalism is a system. It's a state or a condition that a system might promote, might have that property. Just like before banks were invented, you might say, "Well, you could set up a capitalist society," but all sorts of things that the king used to do aren't really possible because somebody who doesn't have the money to start up a business to do the thing can't start it up. And so it won't be started up, whereas the king could just order it to be started up and so on. I don't know if that's a real example, but institutions have nontrivial knowledge in them. There was a time when people invented what we now think of as a bank, and there was a time when people invented what we now think of as money.

And the institutions that we have involving limited government and liberalism and elections and so on, all these details came about as solutions to problems that came up because they didn't have them before. So freedom of speech and so on. It wasn't instituted all in one go. It was instituted in bits, and each bit was thought to solve a problem. And sometimes there were false steps, where people thought something would solve a problem and then it didn't and so on. And what we call modernity or liberal democracy and so on, the features of it, like the banks and so on, they are features that were installed to solve problems. And if you went back to prehistoric times and tried to found a bank, you couldn't because the problem that a bank solves hadn't arisen. You can't have an accounting system among people who haven't yet invented numbers or voting for that [matter]. If you can't count the votes, I suppose you could tally them. But even tallying was invented.

So anarcho-capitalism is a state where no monopoly of violence is considered legitimate. But that can only exist when the problems of not having a monopoly of violence have been solved somehow by some institutions. And so people can be in favor of those institutions rather than…you can't be in favor of just the outcome without specifying the institutions that would cause it. We have institutions that make it unthinkable that mass violence would erupt over some political issue—at least, in many countries that's the case. In other countries, it's not the case, and nobody can put their finger on exactly what it is about existing institutions that have that property. If you abolished existing institutions, then there would be violent conflicts among the force users. And well, I think in real life what would happen is that they would want to revert to the present system. You know, they would decide that the present system is better than the warfare they're having.

Anarcho-capitalists, I've heard them say that "No, they would sit down and they would come to an agreement about how to resolve such disagreements about violence. One company and their customers think that there should be copyright laws enforced violently if necessary, and then the others think there shouldn't be and that should be enforced violently. And then they should sit down with each other and determine who would win a war if they fought one and then consent. The losers would consent to losing the argument without ever having to fight war."

Now, there are no such institutions. Nobody has ever done this. Nobody who was about to have a war has ever opted to not have it because they're going to lose. The issue of whether they need to lose was considered long ago when they declared war, in fact, long before that, too.

So, I do think that anarcho-capitalism points to some real problems with existing methods of existing institutions of consent. They sometimes violate consent unnecessarily and greatly. It is not obvious what to do instead, and it can't be done by fiat and it can't be done instantly. What I would advocate is, instead of trying to create an overarching system which would automatically solve all those problems, I would rather address the problems and let the overarching system evolve from addressing the problems. So there was a time when, in Britain, it was a consensus theory adopted by the majority that the government should seize control of the commanding heights of the economy, as they put it. And so that was done. And it failed in its own terms. And now nobody wants to seize control of the commanding heights of the economy. So, fortunately, this happened in Britain. Because it happened in Britain, it didn't interfere with the memes controlling the use of violence. So the government was able to institute all these stupid things without facing significant violence in opposition. And then, when it was undone later, there was, again, not significant violence opposing undoing it. This is unusual. Usually when a state-based economic system is imposed in a political culture, it's done violently and it encounters violent opposition. And then, well, either it fails or it succeeds. If it succeeds, it succeeds by wiping out the opposition militarily. Obviously, that's an inefficient way of running things, and it's also a way that is antagonistic to the growth of knowledge.

So I wouldn't want to do that for anarcho-capitalism. I would rather see something which, with the[benefit of hindsight, we will say, "Ah, yes, that was what we have now. This system is definitely better than what we had before. And important features of it were foreseen by David Friedman. But there are these other features of it, which are the things that make it work, which weren't always foreseen by David Friedman, except in very general terms."

39:45 **Arjun Khemani** Right. So we need to take it incrementally [and] step by step, solving problems, relevant problems along the way and kind of foreseeing it as a system, like playing out. I guess you could foresee some of the problems and propose solutions to them. But I guess the point about the growth of knowledge being unpredictable does apply here. That you can't even foresee what future humans will want or what future you will want. So, yeah, I think that's an interesting point to keep in mind.

Well, David, I could ask you questions all day, but I think that's a good place to end our conversation for now. Thank you so much for joining me.

40:35 **David Deutsch** It's been fun as always.

PETER BOGHOSSIAN:

FERMI PARADOX, IDEOLOGICAL CONTAGION, AND MORE (WITH REID NICEWONDER)

About the interviewers:
Reid Nicewonder is President of Street Epistemology International and the creator and host of "Cordial Curiosity," a YouTube channel that posts conversations about a variety of polarized topics. Reid follows a variation of the Socratic Method to expose and understand the ways in which everyday people come to have their strongly held beliefs. A graduate from UCF film school, Reid is a cinephile and an avid science & tech enthusiast.

Peter Boghossian is currently a Founding Faculty Fellow at the University of Austin and the Director of National Progress Alliance. He was a Councilman for the State of Oregon (LSTA), the Chairperson of the Prison Advisory Committee for Columbia River Correctional Institution, wrote national philosophy curricula for the University of Phoenix, a research fellow for the National Center for Teaching and Learning, a full-time faculty member in the department of philosophy at Portland State University, an Affiliate Research Assistant Professor at Oregon Health Sciences University in the Department of General Internal Medicine, an advisor for Counterweight, a Senior Fellow at Hungary's Mathias Corvinus Collegium, a national speaker for the Center for Inquiry and an international speaker for the Richard Dawkins Foundation for Reason and Science.

Websites:
https://cordialcuriosity.substack.com/, https://peterboghossian.com/

YouTube channel: https://www.youtube.com/@drpeterboghossian

About the organization: Street Epistemology International is a 501(c)(3) non-profit organization that creating the *Navigating Beliefs* course and supports the development, practice, and promotion of Street Epistemology to people around the world.

Organization homepage:
https://www.streetepistemologyinternational.org/

EPISODE DETAILS

Date	March 5, 2024
Interviewer	Peter Boghossian, & Reid Nicewonder
Source	YouTube
Show	Conversations with Peter Boghossian
Episode	David Deutsch, Peter Boghossian, & Reid Nicewonder Talk Fermi Paradox, Ideological Contagion, & More
Description	Peter Boghossian and Reid Nicewonder, President of Street Epistemology International, are joined by one of their intellectual heroes, David Deutsch. The conversation begins with an exploration of the contagious nature of ideologies, where good intentions can lead to intellectual pitfalls. The three discuss the challenges posed by Social Justice ideology, scrutinizing its evasion tactics in the face of critique and the broader societal implications, such as corrupting peer-reviewed literature and hindering open discourse. Street Epistemology principles underscore the conversation as an antidote to hindering discourse. Other topics in this discussion include problem-solving, philosophy, the Fermi Paradox, cosmic exploration, and more!
Link	https://www.youtube.com/watch?v=qPvTbftn-xY

Ideas

- The society of the West is not one that makes the right decisions all the time. It's the society that corrects errors. Many societies are the opposite. It's not that they're bad at correcting errors. It's that they want to entrench errors.
- One of the epistemological mistakes people make is to assume that we start with evidence and then extract a conclusion from the evidence which is more justified than if we didn't do that. That is wrong. What actually happens is that we start with the problem. The problem is the thing that gives you the coordinates by which to judge ideas.
- Reason is entirely critical. We always find errors and correct them, not find a way to justify an idea as some ultimate truth. That's why it has to begin with a problem. So if we have a problem, a conflict, then we know that either one side is wrong, the other side is wrong, or they're both wrong. A problem is a fertile thing.

Topics

bad explanations • black holes • children • coercion • conjecture and criticism • deriving things • destroying the means of error correction • educational system • empiricist • epistemology • error correction • evidence • explanation • fallibilism • falsifiability • infinite regress • laws of physics • moral realism • objective morality • objective truth • optimism • Popper • Popperian scheme of things • problems • progress • prophecy is impossible • rival theories • Socrates • the Fermi paradox • the West • Wittgenstein • Wokeism

Episode title and description reproduced verbatim from the original source.

Transcript

0:48 **Peter Boghossian** Welcome to Conversations with Peter Boghossian. Today, we interview physicist and author and one of our intellectual heroes, David Deutsch, the Oxford physicist. And Reid and I did this interview together, and it was a very, very special treat for us. Cool. So David Deutsch, Professor Deutsch, thank you so much for coming on to Conversations with Peter Boghossian, and we're here with Reid Nicewonder, the President of Street Epistemology International, and we're big, big, big, huge fans. So excuse us if we fanboy out a little bit about your work. We're big fans. Go ahead, Reid.

1:20 **Reid Nicewonder** Yes. Welcome, David. Thank you. I appreciate you coming on. So, yeah, I know both of you are big fans of science fiction. Anything that you've been watching or reading lately that have been enjoyable to you?

1:34 **Peter Boghossian** I like Severance, I thought that was great.

1:37 **Reid Nicewonder** Severance. Yeah. And I like Devs from FX.

1:40 **Peter Boghossian** Oh, that's great. That's been out a while now. I thought that was fantastic.

1:42 **Reid Nicewonder** That was a while ago. Yeah. What about you, David?

1:45 **David Deutsch** So the last thing in terms of movies that I really liked was The Martian. Because that is not only hard science fiction, which is the only kind of science fiction I like. I mean, the only kind I like qua science fiction. I can watch a silly movie and enjoy it. But as science fiction, not only was it hard science fiction, it is optimistic. It's optimistic in exactly the right way—that it is about problem solving. And it's about someone who thinks that the problem is soluble, and what it takes to solve it is creative thought and—

2:33 **Peter Boghossian** Reason.

David Deutsch Reason, exactly. And if he's not going to solve it, it's not that it was insoluble. It's just that it was bad luck, and he failed to solve it. And if he had his time again, maybe he would.

2:49 **Reid Nicewonder** Yep. I loved that book. That was a big page turner for me. It's just always enjoying the next problem that he was now facing due to having solved potentially a previous problem. So it just kind of builds on itself, but it was just so exciting. I loved it.

3:04 **David Deutsch** Yeah. And in that way, it's quite realistic. That's what really [happens].

3:09 **Reid Nicewonder** Yeah. Exactly. Right. So I don't know how much, David, [you'd] know of Peter, or how much Peter really you know of David. I've just been enjoying David's stuff. I read his book, *The Beginning of Infinity*, last summer, and I've been pretty much enamored with all the ideas from that. But I just want to basically introduce you guys and fanboy out while I potentially help you have a conversation together and talk about some of your ideas and how they potentially compare and contrast so we'll have a good conversation. So, David, I know you mentioned that you enjoyed Peter kind of attacking some postmodern stuff. Do you want to go into that a little bit more?

3:50 **David Deutsch** Well, I think it's very important. At the moment, the postmodern attack on civilization is probably the most dangerous of the intellectual attacks. I can't quantify how much danger there is of nuclear war at the moment. Obviously, that would be worse if it happened. But apart from physical danger, I think…[and] in the long run, the intellectual danger is probably worse because if there's nothing to defend, then it doesn't matter if we give in to nuclear blackmail or whatever.

4:40 **David Deutsch** So I thought that, I don't know what to call it exactly, that experiment or that demonstration was admirable.

4:51 **Reid Nicewonder** The Grievance Studies Affair.

4:53 **David Deutsch** Yeah, yeah. And the attacks on it are just [an] indication of how bad things have got. If I remember correctly, it was attacked for violating the university's rules, you know, just completely pulled out of a hat.

5:23 **Reid Nicewonder** You can't do the testing or have the human subjects be a part of the experiment by submitting papers to journals, that's a human subjects violation.

5:35 **David Deutsch** Yeah, yeah.

5:36 **Peter Boghossian** Yeah. It was pretty crazy. I think in retrospect, part of the problem is that it was too early. If the grievance study stuff had come out a little later, people thought that we had lost our minds, that we were crazy, that we're some cranks screaming about a problem that doesn't exist. So I think that now people are starting to really wake up, to borrow a turn of phrase, to the problem of organizational institutional capture. Activist disciplines, disciplines that subjugate truth to narrative, or—this is what you've written about in your work—the importance of objectivity and using the tools of reason and science to make discerning judgments about things. So I think that people are starting to wake up to the problems and it's going to be a slow burn, but I think we're going to get there. I'm ultimately optimistic that we will get there, but it's going to take a while.

6:40 **Reid Nicewonder** Any comments on that, David?

6:43 **David Deutsch** Well, I don't like to prophesy, but I also think that the ingredients for getting there are there. Our society, [the] society of the West, quintessentially it's not the society that makes the right decisions all the time. It's the society that corrects errors. That's what the West is good at. And I have pointed out from time to time that the optimistic view and the error correcting view and the Popperian view—that prophecy is impossible and everybody's fallible and all those things—imply that there's no limit to the size of error that we can make. The individuals or groups or society as a whole can make arbitrarily large errors. And the important thing then, the important difference between different societies and different groups and different subcultures, is therefore how good they are at correcting errors.

Of course, many societies are the opposite. It's not that they're bad at correcting errors. It's that they want to entrench errors. They don't recognize that there are going to be errors in everything they do, and, therefore, they want to entrench everything they do. And that is their weakness. And that is why, in my view, we have the tools already to win the battle if we use them. And at the moment, it's sort of unfashionable to win anything, and to criticize anything. But, yes. I agree. I agree with you. I think that we've got what it takes.

8:56 **Peter Boghossian** There does seem to be a relationship between the deeper, more entrenched the ideology, the more demeaning of the error correction mechanism.

9:09 **Reid Nicewonder** Perhaps.

9:12 **David Deutsch** Yes. That's the saving grace.

9:14 **Reid Nicewonder** Yeah. And one project I know Peter pioneered back in 2013, potentially to help solve this problem of the lack of even wanting to seek errors for certain ideas, was this project of Street Epistemology. I think that's what you wanted to call your first book, the name of it, "Street Epistemology." I wanted to call it, after reading it, I thought that a best title would be, "A Manual for Creating Critical Thinkers." But it's basically about Socratic method style of having conversations, civil conversations, asking questions to help people critically reflect on the reliability of the methods they use to come to knowledge. And since 2013, tons of people have been taking those ideas and practicing them in the real world, and they have evolved through trial and error, as these ideas seem to do. And you wrote a second book, *How to Have Impossible Conversations*, in 2019. I helped cofound a nonprofit, Street Epistemology International, to help people learn about this way of having more civil conversations, and I want to potentially do more in that arena. Is there anything else I missed about that project, Peter?

10:26 **Peter Boghossian** No. But I'm curious, David, what you think about this. One of the things that I have learned from trying to have these impossible conversations with people, it gets back to your idea of error correction. I think one of the reasons that [people] don't seek out an error corrective mechanism, whether it's as an individual or institutions don't seek it out, is because it's imbued with some kind of a moral virtue. Like, the content of the belief has a moral valence to it in the way that…if you ask somebody a neutral question, they'll adjudicate, like the circumference of the length of a hot dog bun from one restaurant or the other. They'll just pull out a tape measure or what have you and measure the size of the bun because there's no moral component to that. And so I think that there is a relationship between the disposition to seek out errors and one's feeling that they shouldn't criticize certain ideas on moral grounds.

11:43 **David Deutsch** Yes. In a free society, this boils down to, on the individual level, what kind of hang-ups people have. And on the social level, it's about what kind of interaction an individual has with their group of acquaintances, colleagues, other people. If you're in a position where if you say a certain thing, you're going to lose all your friends, then that is a major life crisis.

Peter Boghossian We know that well.

David Deutsch And you have to be brave, and whenever I admit that I am a complete coward in this respect, I have to also say that it's not just cowardice. Maybe I'm just being defensive here, but it's not just cowardice. People have lives. They have something that they want to do other than win this war. Even though, if the war is lost, then everybody will suffer. But on the other hand, if there's nothing to defend, if people don't have lives that they don't want to go away, then it's not worth defending it. So you have to pick soluble problems. And I do what I can, but there's a lot…no. I don't do what I can. I don't do what I could. I do what is convenient, which is not zero. And you do an awful lot, which is perhaps less than 100 percent, but close. So that's very admirable. But I think that physics won't do itself. Epistemology won't do itself, and somebody has to drive knowledge forward in whatever respect interests them most. That was a long apology. Sorry.

14:18 **Peter Boghossian** No, it's great.

14:19 **Reid Nicewonder** No worries. Yeah. I love that. Basically, we want to have a cultural movement to help more people be able to think through ideas in a more critical fashion with critical thinking so that they can potentially solve their own personal problems, but potentially apply them in their own professional lives. Who knows where it could spread. But ideally, one of the big goals is to help maintain relationships at the very least so that these contentious issues aren't driving people apart in terms of their relationship. If someone hears an idea and they disagree strongly, they don't throw facts and evidence and debate and start calling people names. They can have a civil conversation and still actually seek to understand and learn. Potentially, they're mistaken about something. Hopefully, do this [in] a more collaborative civil way, but still doing it rigorously.

15:12 **David Deutsch** You're doing this in practice. I immediately, I would think of the theoretical problem with doing that, which is that the current enemy, which we've been calling postmodernism, but you could call it Woke. One of its features is it keeps changing its name. It keeps changing its identity. So one of the things it immediately does is that when you try and criticize it, when you criticize an idea, the person will say, "Oh, well, that's not my idea." That's just a thing that you are calling it, but our actual idea is something else. And so you never get to substance.

But really, the problem with that whole spectrum of ideologies or single ideology under lots of different names is that it doesn't argue. And its ideology is that that is the right thing to do. The ideology is that you should not argue, because if you're trying to argue, then you are playing the oppressor's game or, you know, playing the patriarchy's game or the white people's game or the Jews' game or whoever it is. And that immediately, in terms of the debate, that's ad hominem, and ad hominem breeds ad hominem. So the discussion is automatically taken away from the substance of the discussion and onto issues like, "Am I a bad person?" And in terms of the opponent, the other speaker, it has an even worse effect on them because they are confirmed. Every time this happens, they are confirmed in their ideology that one should try even harder not to argue and to base everything on ad hominem chunks of predigested pseudo-arguments like, "You were only saying that because you're a whatever."

17:42 **Peter Boghossian** Yeah, yeah. It's very interesting to me. I think if you look at it as memetic, if you just look at it as a kind of idea pathogen or a kind of almost a contagion, it has developed defense mechanisms to keep it in place. And chief among those defense mechanisms are, "You do not talk to people who have substantive disagreements with you. You do not talk to people who will challenge or question you." And again, it's playing on that moral idea. And doing so makes you a better person because you don't want to platform people who have different ideas. You don't want to give those ideas the opportunity to spread. But what that really is, is an excuse for the person who refuses to engage in the conversation not to do the intellectual work. But even more so, it's a kind of sheath that prevents the ideas themselves from the potential of being revised.

18:48 **David Deutsch** Exactly. And I used to think that and maybe I still think that, this kind of mental self-defense mechanism to protect the bad ideas is caused by psychological hang-ups, which are themselves caused in childhood by parents. And that is a different idea from the idea that it's a mind virus. When I say by parents, I mean by parental coercion, by intentional actions. Perhaps not consciously intended to have that effect, but it's the fact of coercion that does the damage. I used to think, or I still think. I don't know. I'm not sure. But with the phenomenon that looks like a mind virus, we have people entering into the ideology or into the infected state somehow voluntarily, joyfully, and with good intentions, and never thinking that they are limiting themselves. Thinking actually that they are expanding their consciousness, they're waking up to become Woke. And apparently it feels subjectively like coming to understand a real thing, coming to understand something, and that you didn't understand before.

So another place where this happens is in cults. I think for a wicked cult to be effective, it must use some kind of coercion.

But it may only be the coercion of, "Oh, I'll lose my friends if I even think something taboo." Because, as you said, engaging in an argument with a racist or a whatever it is means you are bad. It's not just that they are bad. It's that you are bad, and you don't want it. You keep trying to think the opposite way to that. And that every time you do that, it takes you deeper into the cult. But Woke doesn't have compounds where you're not allowed to sleep, and it doesn't have that mechanism of inducting people into a adult onset hang-up. So I don't know exactly how it works. I don't know why it works. You probably have sophisticated theories of how it works. So tell me if you do.

22:02 **Peter Boghossian** I do. That's a conversation for another day, but I definitely think it starts with the wide-scale capture of not only academic institutions, but colleges of education where they grant degrees or licensures, depending on the country or what have you, certificates to teach in K through twelve, and then it's a massive indoctrination mill. It's kind of an ideological replication factory.

22:28 **Reid Nicewonder** Right. And one of the features or qualities of Street Epistemology is that it has the ability to tackle ideas or beliefs that are largely maybe driven mostly by psychological or social motivations. Some kind of these beliefs might meet very deep needs for people, and we still try to get at those needs in terms of talking about potential biases. And if they come up as potential reasons for thinking something is true, we can talk about it in that way. But ideally, keeping it in the frame of, "Okay, is this really solving a problem for you? Is this the best way to solve this kind of problem for you?" It's almost maybe more therapy than an epistemology conversation, but it's back and forth.

23:11 **David Deutsch** That sounds like a very good kind of attitude to take to get people out of the hang-up, the cult, or whatever you call it, the mind virus. But how [does] one [get] into it? So I'm very skeptical that school teachers and university professors are the origin of the problem. I would more likely think that they are just another symptom, because I don't see schools, for example, being very effective at teaching anything. If you think that nearly everybody at age whatever it is, fourteen, can solve a pair of simultaneous equations or draw a graph of y equals x squared or whatever. And yet, if [we] ask them, when they're double that age, most of them won't know. And that's a sign of the ineffectiveness of teaching as we now know it, the ineffectiveness of teaching a thing that the person didn't really want to learn in the first place and has no use for. So why is it that an entire generation can be taught nonsense and behave as if it were fundamental to their life? They would say, "This is what I am as a person." Now, if schools can't do that for algebra, for science, for English, all the things that they think they are teaching, how come they can do it for Wokeism? So that's why I'm reluctant to accept that explanation of how the thing began.

25:30 **Reid Nicewonder** I know you have your substitution hypothesis where maybe certain traditional religions have been falling away, and some of these postmodern ideas maybe meet a lot of the needs that traditional religions have been meeting, and these have just been now expressed in, potentially, academic settings, and then that gets filtered through the culture. Hard to know.

25:53 **Peter Boghossian** Right. And it's reinforced by community, friendships, people who believe these things, administrators and faculty who get, for example, diversity statements in hiring, faculty who get hired and promoted on the basis of non-meritocratic criteria. And so you're really creating a system where you're aligning things, but to me, all of that, it's not merely that it begins in the university and is perpetuated and forwarded by the university and colleges of education in particular. But friends, groups, are then bundled within the ideological framework.

The other part to that is the peer-reviewed literature, is the corruption in the peer-reviewed literature and the fact that these ideas are really from activist disciplines that are being forwarded. [They] don't have truth at its core. They're not falsifiable. They're not empirical. They're really the musings of ideologues discharged in peer-reviewed journals. And so I would agree with you in that it is not singularly the fact of the university culture. There are so many other elements that go into it. And not only university culture, but in United States, you call it K through twelve, kindergarten through twelve culture.

27:22 **David Deutsch** Yeah. But people in the educational system, it's notorious that the thing is set up as a machine to pass exams. You know that you have to act and speak and write in a certain way to pass the exams. And then you just shake it off like a dog coming out of a river. [Most people] don't really internalize. There are people who are actually interested in each subject, but those people are going to be the most critical because they want the ideas for a reason, and they're going to want to improve on them. That's another thing. Woke people don't want to improve on Woke ideas. They just adopt them. It's notorious that somebody can come out of a science course and not know the first thing about the subject, about the science in question. They just know what to do to answer the questions because they're not interested. And therefore that's a bad thing when it's a good subject, but it's a good thing when it's a bad subject, when it's a subject that doesn't really exist.

Why is it that they just come out of that and say, "Yeah. Well, you know, I had a nice three or four years, drinking with my friends and whatever," and saying that nonsense, "and now I can start my life properly." Why don't they say that about the Woke thing? Well, maybe some do. Maybe most do. That's another thing maybe I'd like to ask you. How many people does it take to have got the mind virus in a certain culture or subculture before the whole culture is disabled? Maybe it's only ten percent, because, you know, maybe it's not necessary that the person thinks, "I'll lose all my friends." Maybe it's enough if you say, "I'll lose ten percent of my friends, and they will really hate me." I don't know. I'm asking. I don't know how this works.

30:01 **Peter Boghossian** I think my response to that is, it's not a percentage of the population or a number, although there does seem to be a minimal critical threshold. But it's a play on the, "All it takes for evil to flourish is for good men to do nothing." All it takes for the city of Portland to burn is [for] people to not do their jobs. Mayor Ted Wheeler—him/him/his—to not do his job. The police to not arrest people. The prosecutor to not prosecute people. The judicious…So there has to be a kind of systems failure. People rip down statues, there's no consequence. People are assaulting people on the street, there's no consequence. Police refusing to go into minority neighborhoods, and consequently the murder rates in those neighborhoods go up, there's no consequence. So it also has to be a breakdown in systems where people who don't subscribe to the ideology are absolutely terrified of a pile-on on social media of being called a bigot, a homophobe, a Nazi, a racist, when the overwhelming majority are clearly none of those things.

And so I think it's a kind of culture of fear that's created. And once you've created a culture of fear, you put a wrench in the machine, the machine stops functioning. So I think a framework for the question is not how many people subscribe to deranged ideas, specifically ideas that want to really destroy the foundations of Western society, but how many people will do nothing in response to that. To be very blunt with you, David, they're just cowards. They're afraid to speak out.

31:46 **David Deutsch** Well, in a totalitarian society, it's obvious what they're afraid of. They're afraid of the midnight knock on the door, and the secret police coming to arrest them. In a Western society, that's not going to happen most of the time. It does happen some of the time, but I think that's not what people are afraid of. They're afraid of other people. And why are they afraid of even speaking? I don't know. Again, I'm speaking from first principles, from sort of dead reckoning. If what you say is true, that most people don't actually subscribe to the idea that they are acting out, why is it that…you say the police don't go out into that neighborhood to arrest people. Okay. Because they're afraid of something.

Why don't, why aren't they then sitting in their common room or whatever police have, drinking their coffee and saying to each other, "Wow, you know, we really should be going out there, but I'm not going to put my whatever it is on the line." And the other one [said], "No, nor am I." You know, so is that what happens? I don't think it is. I think if ninety percent of the people were talking to each other about how awful it is that they're not allowed to say certain things, then from that state, you get automatically to the state where they are saying things in public.

33:29 **Peter Boghossian** I'll push back on that a little bit. So there's something called the Ferguson effect, and it's, basically, police didn't want to go into the Ferguson neighborhood and arrest, specifically, African Americans, because they were afraid that if they did, [that] they would be deemed as racist. And the consequence of the lack of police presence in those neighborhoods, it was devastating to the black community, specifically young black men who kill other black men. So I do think that there's a kind of fear created both within institutions and external to institutions around…it could be around anything, obviously. Any kind of totalitarian state can arbitrarily pick it, but at the moment, it's around race, primarily race. It's also, to a certain extent, around trans status. Andrew Doyle writes about this in his book, *The New Puritans*, where men are in women's spaces and people are afraid to say anything because they'll be deemed bigots. So I think understanding this phenomenon is vital to understand the culture of fear that's created. And really, when you think about it, it's so idiotic. There are only a few tools in the tool set. Nazi, bigot, homophobe, racist, sexist, misogynist.

34:54 **Reid Nicewonder** Yeah. And hopefully, one potential solution to this problem of not being able to talk about things is to have a really great way of having conversations…Street Epistemology. But a core part of Street Epistemology is the epistemology, and I want to make sure, are we actually using the best theory of epistemology to have these conversations with people? And so I [want] to potentially transition to that to just ask this question. Does the following sentence make sense? "It is rational to align one's confidence to the evidence." I know you (Peter) want to expound on that a little bit more.

35:31 **Peter Boghossian** No, I think that's good.

35:31 **Reid Nicewonder** Okay, go ahead.

35:33 **David Deutsch** No. But let me say that differences between people in regard to the foundations of epistemology, that is, what knowledge is, where it comes from, how it grows, that kind of question, foundational questions. I think most people, most philosophers, as you know I'm a follower of Karl Popper, and that is a minority view on every level among philosophers, scientists, people in the street. The prevailing epistemology is some kind of inductivism, empiricism, and, perhaps worse, positivism, instrumentalism, those kind of things. And those are, as foundational theories, those are really, really bad. But the saving grace is that foundational disagreements usually don't strongly affect the conclusions that people draw. So you can say, "I'm against Woke because," and then the Woke person will say, "No, no." But never mind that. Because, they don't, whatever you just said, "align themselves to the evidence."

Now, I would say that they try to destroy the means of correcting errors. They don't judge ideas by their content but rather [by] their source. And all sorts of Popperian maxims like that, which conflict with the empiricist ones. When it comes to actual cases, we will 99 percent agree about whose side we are on in particular cases. We'll just explain it differently. But having said that, let me say that there are some things, some cases where having bad epistemology takes you way off course and often into the enemy's camp. So I do advocate getting the right epistemology.

38:06 **Peter Boghossian** And what do you mean when you say 'the enemy's camp'? You mean the conclusion will be arbitrary?

38:12 **David Deutsch** Well, with some of these epistemologies, the conclusion will lead you to endorse the enemy's point of view. For example, there's this famous debate between when Wittgenstein and Popper met, and Wittgenstein allegedly threatened Popper with the poker. And the issue that he was threatening him with was that, Wittgenstein had said, "There's no such thing as a philosophical problem. There are only language problems. And there's problems with the way we use words and the way we define concepts and that sort of thing." And Popper said, "No. That there are genuine philosophical problems, and it's possible to solve them, and it's possible to make progress with them if we don't solve them." And Wittgenstein said contemptuously, "Give me an example of a genuine problem…a genuine moral problem." I'm not telling this story very well. Somebody wrote a whole book about this exchange, by the way. And Popper said, "An example is not to threaten visiting speakers with a poker." And, supposedly, Wittgenstein stormed out of the room.

Now the thing is, if you adopt Wittgenstein's view there, then you will have no philosophical weapons against the argument that, if there are no philosophical problems, then there's no such thing as objective truth in things like morality. So you will think that you can't say objectively that slavery is bad. You can only say a certain culture thinks it's bad and another culture doesn't think it's bad. And if the worst thing in the world is for one culture to oppress another, then you can't oppose slavery across the board. Or, in the most recent example of this, you can't oppose rape across the board or kidnapping or torture or murder. So there's a case where adopting the wrong epistemology ties your hands in regard to arguing against really the worst possible conclusions.

Let me just say, again, that this is the exceptional case. In most cases, it doesn't matter whether you begin with Wittgenstein or Popper or Bertrand Russell or whatever. All these things are attempts to codify common sense and to improve common sense where necessary. And so all those bad philosophies began like that. And to some extent, they embody still common sense in their ways of arguing and so on. And if you're lucky enough that the thing you're arguing about can stay on that ground, then it doesn't matter that you began with the wrong epistemology.

41:44 **Reid Nicewonder** Great. Yeah. Maybe one last detour on [this] epistemology, and especially when it relates to ethics. I know, Peter, you're a fan of saying, maybe if you want to justify why slavery is bad, you would say, "Well, ultimately, it's rationally derivable."

41:55 **Peter Boghossian** Correct.

41:56 **Reid Nicewonder** Yeah. What would you say about that, David?

42:00 **David Deutsch** Again, as a fan of Popper, I don't think we derive things. I think we, speaking about any field, but let's say morality in particular, we have certain ideas about what we should do, what other people should do, how to think about other people, what to expect from other people. And these ideas are imperfect. They have problems in them. You have the idea that murder is wrong, and then you come across a problem of: somebody who kills his wife because she's suffering unbearable pain in a terminal disease. And is that wrong? Is that murder? So we come across problems, and the reason why there is such a field as ethics is that there's knowledge to be obtained about how to solve such problems. It's never perfect.

43:11 **Peter Boghossian** So when you say there's knowledge, you're saying that there's a fact of the matter to be figured out.

43:18 **David Deutsch** Yes.

43:19 **Peter Boghossian** That's what knowledge means.

43:21 **David Deutsch** Yes. There are facts of the matter. There are objective truths. We can never grasp an objective truth completely. Our ideas about it will always be flawed in some way. But if we can make progress enough to solve the problems we have, then we're making progress. That's what progress is.

43:43 **Peter Boghossian** Right. So if there is a fact of the matter about things or if there's knowledge, then the question is: Are there better and worse ways to figure that out? And the answer, the moment that you say, "Yes, there are worse ways," that must, by definition, mean there are better ways.

44:03 **David Deutsch** Yes.

44:05 **Peter Boghossian** So sacrificing a goat on the hood of your car to figure out something is a worse way, so that must mean that there's a better way. If you just escalate that up, that idea, science, evidence, reason...Well, reason, you could use reason and then take data points, take evidence to inform your decision. But again, that's assuming that there are facts of the matter to be known, like moral facts, and then you just have to use reason to get there.

44:42 **David Deutsch** Again, I should keep mentioning Popper because it's really him that I'm trying to summarize. But one of the mistakes of epistemology is to assume that we start with evidence, or that we should start with evidence, or that if we want to make progress, we need to gather the evidence and then think about the evidence, and then extract from the evidence, a conclusion, which is then more justified, more secure, more firm than if you didn't do that. And that's the wrong picture. As I keep saying, you may not go far wrong if you follow that picture, but it is completely the wrong picture. What happens is that we start with the problem. The problem is the thing that gives you the coordinates by which to judge ideas.

So a problem is a conflict between existing ideas. So, somebody says, like a few years ago somebody said, "We've discovered in our neutrino experiment that the neutrinos are traveling faster than light." And the press was saying, "Oh, you know, maybe Einstein's theory of relativity is wrong." That was a mistake. That's the empiricist mistake. The meaning of the evidence is not visible until you have at least two rival theories which both purport to explain it. So you can't just say Einstein's theory of relativity is wrong because that's not an explanation of anything. Scientific theories are explanations. We have explanation in Einstein's theory. To say that Einstein's theory isn't true is not an explanation. The negation of an explanation is never an explanation.

So you have a rival explanation which says that neutrinos sometimes travel faster than light because x, y, and z, and there wasn't such a thing. So people looked for such theories, but they also looked for theories along the lines of, "Maybe the apparatus isn't doing what we think it's doing." Both those things are lines of research trying to find a rival theory. It's only when you have the theory and the rival theory that evidence is even meaningful. Once you have the two rival theories, then you can gather evidence because then the two rival theories will be giving meaning to that evidence. If you only have one theory or if you have zero theories, then the evidence is meaningless.

47:47 **Peter Boghossian** Yeah. I have a comment on that. Or do you want to move on to another thing?

47:51 **Reid Nicewonder** Yeah. Sure. Quick comment.

47:52 **Peter Boghossian** So my quick comment is: I agree, and this is going to be a weird criticism of what you just said, but with the exception of a time index. So if you're in kind of a hurry to figure out which rival theory one should consider, or more specifically what one ought to do morally, then you would need to start with evidence. So let me give you an example. It's such a jarring example, and I apologize. In South Africa, it is a common belief that if you have AIDS and you rape a baby, it will cure your AIDS. So before you even go to the idea of rape, if you started with your evidence that that's false, then you could just a priori rule out the whether or not one should do that, whether or not one should engage in that behavior. So it would be a way of like a gold sieve when you drop in evidence, it would be a way of being able to immediately discount a moral conclusion.

49:11 **David Deutsch** Well, I think you can do that even without evidence, because what you have there is two rival theories already in the problem as you described it. There are people who have this false theory about AIDS and babies, and there are people who have a different theory about what AIDS is, and also moral theories about babies, and so on. So there's a whole spectrum of theories in conflict with each other. Now, you could resolve that conflict with evidence, but I don't think you need to do that because in order to say that the quote "scientific" side, the scientific theory is false, you need to have an explanatory theory of what is true. And in this particular case, and in almost all of the anti-science kind of worldviews out there, you can't sustain an explanation of why the scientific take on the issue in question is false unless you invoke a conspiracy theory. And once the other side is invoking a conspiracy theory, they won't accept your evidence. They will just say, "Well, they're just lying." Therefore, in a case like that, argument is going to solve this. Evidence isn't going to solve it. You can then say, "Well, these people who are faking the evidence, what happens when they get AIDS? What happens if their child gets AIDS? Do you think that they take the drugs that they advocate, which you say they know very well don't work, and that really the rape works? Which do you think they do?" And they will say, "Well, yeah. Then they secretly rape." Then you say, "And what do they tell their friends? What do they tell their children? At what age do they reveal to their children that they've been lying about their life's work? How do they make sure that children don't tell their friends in school?"

51:42 **Peter Boghossian** Can I ask a follow-up question, Reid?

51:43 **Reid Nicewonder** Sure.

51:44 **Peter Boghossian** So follow-up question and then we'll go on with the other question.

51:50 **Reid Nicewonder** These are the kind of examples you get with talking with Peter, so enjoy.

51:54 **Peter Boghossian** Yeah. My question to you is, I'm going to keep pushing back on this evidence thing if that's okay. So I'll reveal my cards here. I think one of the problems that I have with philosophy writ large is the idea that you can reason your way to certain conclusions, particularly certain conclusions about phenomenon or phenomena. For example, you can reason your way, and you see this in theology a lot, theologians who are philosophically adept, they can reason their way to the origin of the universe, for example. I think that there is no reasoning your way to the origin of the universe. You have to start with some kind of evidence for that, and I'll just throw this out.

We can change this as a placeholder, but I am fascinated by this idea, and I actually asked one of my professors this and he literally yelled at me when I asked him this: I don't understand why there can't be an infinite regress. I've never understood that. But that would be an example of a thing that you couldn't reason your way to that. Right? So don't you, at some fundamental level, either in attempting to adjudicate between theories or you would need evidence at some level.

53:43 **David Deutsch** I think that this picture of what reason even is, is wrong. It's not a way of getting to a conclusion or of justifying a conclusion or coming from some secure thing like evidence to reach the conclusion. None of those is true. Reason is entirely critical. Which is why, it's always finding errors and correcting them, not finding a way to justify a thing and say that that's now finally the truth. That's why it has to begin with a problem. So if we have a problem, a conflict, then we know that either one side is wrong or the other side is wrong or they're both wrong, but they can't be both right. So the problem is a fertile thing in that sense because…and then you can begin to think about the problem. And [what] you're supposed to think about in the Popperian scheme of things is, "Well, if one of them is wrong, what's wrong with it?" Try to find a thing that's wrong with it, which means criticize it. Conjecture, criticisms. "Well, it might be that the scientists are all lying. Okay."

55:23 **Peter Boghossian** But the criticism of, for example, that there can't be an infinite regress would be based upon a biological understanding of reality that we have. Richard Dawkins calls it the Middle Kingdom, in which our brains have evolved to understand the medium-sized dry goods, I can't remember the exact phrase off [the] top of my head. I was just listening to a podcast the other day, Michael Shermer's most recent podcast, [and] this guy was explaining the quantum realm, and it was just so bizarre. It was just so freaking weird. I couldn't take the podcast more than ten minutes. My head hurt trying to think about all these things. But I think my head hurt because I didn't evolve to think about what happens inside of a black hole or the quantum realm. And so the idea of error correction in those senses doesn't really make sense. Right?

56:34 **David Deutsch** The thing is, people came to the theories about black holes and the interior of black holes, not by building them up from something that our brains were designed to do, but by finding errors in the previous theory. And the enterprise of finding those errors found really bad errors that nobody, literally nobody, can find a way of papering over. That's why we think that black holes exist. But the theory of black holes is undoubtedly false in some respects. We just don't know yet in what respects. And when we solve that problem, I think we will have a different conception of black holes, but we'll still think black holes are there. The infinite regress problem that he was saying that we must rule out infinite regresses, that is really mostly relevant in the wrong epistemology. That is, "Is this the foundation? No. Is the thing below it the foundation? Well, why can't we just have an infinite set of foundations?" You know, flat Earth, and below that the six elephants, and below that the turtle, and then turtles all the way down. That's the wrong way of looking at it because there's no problem against which you can judge these turtles.

58:14 **Peter Boghossian** I'm thinking about what you said about error correction. So for example, the physics of black holes, I don't understand Ramanujan's mathematics. I mean, I don't even have the most rudimentary tools to understand how I could even begin to go about correcting something like that. I don't think we're calling math evidence for something. I'm just trying to think of what tool sets…you would need certain tool sets in order to figure out whether or not the domain into which you're inquiring, again to borrow a Popperian thing, you could have elements of falsifiability. Right? But I think that's different in the moral domain.

59:18 **David Deutsch** Yeah, I agree. Falsifiability is much less important than most people think it is. It's not the cornerstone of Popper's epistemology or anything, and it's only applicable to science. And there are areas of knowledge, including epistemology itself, which are not suitable for applying that tool to criticize theories. If you're wondering how you would criticize a theorem of Ramanujan's, I would ask you, "What's your problem with it?" If someone said, "No. It should be like this rather than like that," how would you choose between them? Well, you're saying you don't have a way of choosing between them. That's because you're not interested in the difference between them. So if you were interested in the difference, you would be able to judge the two different ways of looking at the problem by the criterion of what you wanted the theorem to be like. And in the case of, say, the famous case of the turtles and so on and the Earth being on the top of the turtles, it depends what your problem is. If your problem is what holds the Earth up, then it's no good having that infinite stream of turtles because the whole stream could fall down whether it's infinite or not. Therefore, postulating an infinite number of turtles doesn't address the question you asked, the problem you had.

The problem you had was what keeps the Earth up, this flat Earth. Why doesn't the whole flat Earth just fall down? Then that is simply not answered by the turtles theory. And that is why the infinite number of turtles is a bad explanation. Not that infinity is by itself unacceptable. In fact, infinity is often acceptable. It depends what the problem is and what the criticism was and whether the infinite number of whatever it is meets that criticism or doesn't.

1:01:43 **Reid Nicewonder** Right. And speaking of tools, I have a question about this, and you talk about this a lot, where, say, someone has all of the critical thinking tools and rationality tools that maybe David Deutsch has in terms of finding error correction. Is there something prior to that in terms of one's disposition or attitude or values to even, like, want to seek error? That's a big problem I always try to think about how to solve. Anything to add?

1:02:11 **Peter Boghossian** That's absolutely spot on. Yeah.

1:02:14 **David Deutsch** So we don't know how the mind creates conjecture and criticism. If we knew that, we could write an AGI program right away, and we don't know it. But we do know some things about it. Note that every human who is born and [that] doesn't have brain damage learns their native language. Learning a native language is incredibly difficult. We don't even have the understanding yet of what a language is. Again, that's the problem [of] AGI. But we know that it's a very complicated thing because if we try and do it again when we're adults, we can see how difficult it is. And it's nowhere near as difficult as what a baby's task is because we already have a language. We can connect the different bits of grammar and vocabulary into the structure that we have already built. It's much easier than to build that structure in the first place, which every child does easily. So if [we] think in terms of Popper's epistemology, we don't know how it's done, but we know some things about how it's done. And what we mainly know is that it's done by solving problems. It's done by conjecture and criticism. [So] criticism to try to guess where the mistake is in our theory and conjecture it.

1:04:00 **Peter Boghossian** So everything you just said is a skill set and not a disposition. So don't you have to have the disposition to want to do that in the first place?

1:04:12 **David Deutsch** Yes. Well, I suppose, psychologically, that's what a problem is. A problem is something that you want to resolve. Like, you don't want to resolve Ramanujan's problem, but you do want to resolve the problem of how to defeat Woke, let's say.

1:04:34 **Peter Boghossian** Yeah. I don't have the tools to solve. My math is truly atrocious, but there is a kind of disposition to want to correct your errors, a disposition to not believe that your beliefs are infallible, as crazy as that sounds.

1:05:03 **David Deutsch** Yes. So I think you do need that, and I think we are born with it, as is evidenced by the intellectual performance of babies and children. And I think that what happens—this comes back to what I said earlier in this conversation—what can happen is that this disposition and the relevant mental processes are sabotaged. There's a spanner put into the works so that we lose the ability to do those things. And somehow, it can happen. This can happen even later in life. And that's what I said earlier. I don't really understand how it happens later in life. I think I have some kind of understanding about how it could be sabotaged when a person is young. But once they have learned and understood enough to live a life, even poorly, then how do you sabotage the very means by which they have got whatever they have already got? I don't know how.

1:06:26 **Peter Boghossian** Do you think that not cultivating a disposition to correct errors is a moral problem?

1:06:36 **David Deutsch** Yes. If I understand you correctly. I think that the basic moral problem is what should I do next. Losing the ability to criticize some idea of what you should do next is morally wrong. I'm not sure I understood you correctly.

1:07:01 **Peter Boghossian** No, you did. Yeah.

1:07:04 **Reid Nicewonder** Great. Maybe two more topics. Here's one: Plato and Socrates. I know *The Republic* is one of your favorite books, Peter. And David, you've written a whole chapter on Socrates and dialogue with Plato, and I want to kind of tie that in with the ethics of what Socrates was doing by going out on the street and being a gadfly. Maybe Street Epistemology is not as coercive. We don't really accost people with, like, "Give me your belief." I'm going to ask you questions about it. It's more consensual. We get a lot of informed consent. But the basic ethics of that project of going out in public and helping people think about their ideas, what do you think about that? Is that something even we should be doing in this day and age like Socrates did?

1:07:54 **David Deutsch** Yeah. Well, I think Socrates never forced people to listen to him. They're always free to walk away. What was keeping them talking to him was that he was saying outrageous things, and they wanted to contradict him. And that's good. That's the moral way to attract people to chat to you. That's what does not happen in schools, let's say. Schools do the opposite of that. They say that you have to listen whether you're interested or not. And you have to end up with a predetermined answer whether you're interested or not, whether you agree or not, or whatever.

However, I think Socrates, as portrayed in the Socratic dialogues, is I think unnecessarily abrasive. People are engaging with him despite the fact that he's kind of, I don't know what the right word is, but he's a gadfly. He's stinging them. Whereas I think it would have been better for him to say, "Okay, well, I've been thinking about this thing. How do we know that the sky isn't going to fall down?" And a person might be interested by that. Whereas what he and, I suppose you, now that I come to think of it, [want] to do is that you have in mind a certain error that you think people are making. And when somebody is obviously making that error, you want to challenge them. That is what Socrates did. But to answer your question, I don't see how that can possibly be wrong to do that. I mean, this is the fundamental activity of humans, is to criticize and conjecture. As long as you're not in a position of authority or coercion, you're doing the right thing. The person can walk away.

1:10:28 **Reid Nicewonder Awesome.** Yeah. I'm usually out there in the public with basically this setup with a sign, and I'll let people come to me and we chat about whatever they [bring] up. Nobody's coerced. No coercion as far as I can tell. Perfect. Awesome. Thank you for that. And maybe the last topic I want to talk about: pessimism versus optimism.

1:10:47 **Peter Boghossian** Before we do that, may I ask you one question?

1:10:50 **Reid Nicewonder** Sure.

1:10:51 **Peter Boghossian** Something that's been tormenting me for years. Richard and I did an event on this. This is my first question to him. The Fermi paradox, where is everybody, and is it possible that our model of the universe is completely incorrect?

1:11:12 **David Deutsch** Yes, it depends what you mean by 'completely.'

1:11:18 **Peter Boghossian** Well, that we're in a matrix or a brain of that or something like that?

1:11:20 **David Deutsch** No. That is the epitome of a bad explanation.

1:11:27 **Peter Boghossian** Okay.

1:11:29 **David Deutsch** The brain in the vat or the computer or this computer simulation or in the mind of a demon or whatever. Those are all the same theory, and it's infinitely variable. There's no reason to choose one of them over the other. They're uncriticizable. And an uncriticizable theory is automatically to be rejected because it can't possibly fit into the scheme that corrects errors. So we could always be wrong. It could be that there's the pixies sitting at the bottom of my garden who are going to come into the house and kill me the moment we end this conversation. But I just made that up. I could make up a thousand stories. And it's not rational to make conjectures that aren't attempts to criticize a theory which is part of a problem that you're interested in.

1:12:30 **Peter Boghossian** Okay. So with the Fermi paradox, Carl Sagan said we could be the first. That would explain it. We could be the last. There are other explanations, like a Great Filter, etc. But is this one of the few examples where the lack of evidence indicates that the models that we have are not accurate?

1:12:56 **David Deutsch** I don't think it's such a great problem, basically. It could be that our whole model of cosmology is wrong. And, in fact, our whole model of cosmology was changed about twenty years ago when the accelerated expansion was discovered. So, it doesn't actually affect the Fermi paradox, but that kind of thing could affect it.

1:13:23 **Peter Boghossian** Thank you for not taking offense at my...

1:13:28 **David Deutsch** No. That's the opposite of what I want to do. I think the problem is less problematic than it seems for much more prosaic reasons. There are all sorts of ways in which the universe could be full of intelligent life, and they haven't made contact with us. Just for a start, it could be that they are, and this is acknowledged by the theorists of the Fermi paradox. It could be that they're not interested because they have already made contact a million times with primitive civilizations, and it's no longer interesting enough to be worth sending out probes across the universe and having a conversation at 10,000 years between sentences. That's just boring. So that's just one simple reason.

Another one is that they might be expanding downwards. Like Feynman said, "There's plenty of room at the bottom," by which he meant that, in terms of orders of magnitude, there are more orders of magnitude between us and the smallest possible thing than there are between us and the largest possible thing. And so every time you expand, let's say, to another star, you're going to increase the time it takes…say you have a culture that exists on two stellar systems. Then if it takes, say, four years to the nearest star, if it takes four years there and four years back for the next new idea, then the benefit you get from the other civilization will always be delayed by at least eight years. And in fact, in practice, more than just a back-and-forth will be needed. It's going to take more than just one cycle of back-and-forth. Supposing it takes ten cycles of back-and-forth to become uniformly approving of iPhones. So somebody says, "iPhones would be a good idea," and someone says, "No, it is a terrible idea because you need haptic feedback on your clicks," and so on. So somebody has the idea on one solar system, for that idea to travel to the other solar system might take a century. And meanwhile, the ideas on each planet separately are proceeding on a timescale of years, not centuries. So the conclusion for that is that the aliens might not want to expand in the sense of going to distant places. They might want to expand in the sense of shrinking themselves down so that more and more and more of them can occupy a given volume.

1:16:50 **Peter Boghossian** So one comment and then, Reid, one more question, I promise. That's it. Then I'm going to shut up. So my other question is that, let's just say that I accept by fiat that the overwhelming majority of intelligent, shall we say, spacefaring life beings subscribe to that. Well, given the Drake equation and the possibility [that] the universe could be teeming with life, we should expect to see some that don't subscribe to that, and so we should expect consequently to see something like a von Neumann probe or some evidence of that, but yet, again, we see nothing. So that's just a comment.

1:17:36 **David Deutsch** You don't have to accept anything. There's an argument. There's the Fermi argument that says that it's really weird that we don't see aliens. And that if they were there, we'd see them. What I said to you is not the rival theory. It's a criticism of that argument. It's a criticism of the idea that that is terribly puzzling. And I only mentioned one way it might be okay. There are many other ways it might be okay. And I don't have to believe any of those. In fact, my guess is that the answer is one that nobody's thought of yet. But it doesn't have the aspect of an insoluble problem.

1:18:24 **Peter Boghossian** Okay. My last question is: Do you think that the laws of physics are fixed? Like, if the speed of light is 186,000 miles per second squared, do you think at some level of technology, we would be able to change that?

1:18:45 **David Deutsch** If we can change it, then that wasn't a law of physics. By the way, it is per second, not per second squared.

1:18:52 **Peter Boghossian** Oh, sorry. Yeah. There you go. Yeah. I'm not a physicist.

1:19:02 **David Deutsch** If the laws are changeable, that means that there is some deeper law according to which one can change the subsidiary law. For example, Ohm's law, actually that was never thought to be a law, but let's pretend that it was once thought to be a law, and then people discovered semiconductors. And then they found that we can actually change the behavior of materials so that they don't obey Ohm's law. And the next thing is you discover transistors, and the next thing you discover computers, and so on. But then, we would say with hindsight that Ohm's law actually never was a law. It's just an approximation to a law.

And similarly, the speed of light, we know that the universe itself, space, can be expanding at faster than the speed of light. In fact, distant space, you know, a few billion light years from us, is receding from us faster than the speed of light. Even though nothing in space can exceed the speed of light. So our previous conception of what restriction the law of the constancy of the speed of light imposes on us is slightly wrong. So, in short, if it can be changed, it isn't a law. And in fact, if it can be changed systematically, that means that there's another law according to which we can change it systematically.

1:20:40 **Peter Boghossian** Right.

1:20:42 **Reid Nicewonder** Gotcha. Alright. Last topic.

1:20:46 **David Deutsch** You keep saying that.

1:20:49 **Peter Boghossian** Sorry.

1:20:49 **David Deutsch** No. No. It's great.

1:20:50 **Reid Nicewonder** Peter, I think it's safe to say you're fairly pessimistic about at least the near future.

1:20:56 **Peter Boghossian** Correct.

1:20:57 **Reid Nicewonder** In terms of the West or the United States, mainly, partially Europe. Do you want to make a quick case as to why you're pessimistic in a concise way, and then we'll get David's thoughts?

1:21:09 **Peter Boghossian** Massive deficits, widescale institutional organizational capture, problems with not talking about problems, immigration. I mean, there were just so many problems. I just don't even know where to begin, but suffice it to say that I'm pessimistic about the future of the West. Attacks on liberalism, attacks on Enlightenment values, people, and I don't think that these are merely fashions. I think that there's something fundamental about the values of civilization, science, reason, epistemic adequacy, if you will, or epistemology. I think that truth has been sufficiently demeaned that I don't know how we're going to recover from this. We don't trust our institutions. But anyway, if I were to just pick one variable, it would be: I do not see how we're going to pay back our deficit. I just don't think it's possible. Every economist with whom I've spoken, not a single one has said, "Oh, yes. This is not..." And just in the last month, it went to 34 trillion. So that, coupled with all of the other problems that we face, is not making me a happy camper.

1:22:36 **David Deutsch** Well, I disagree, as you guessed, but let me say first that I don't think there's any law of nature or inevitability or prophecy that says that we must solve this, it's impossible that the West will be destroyed, it is impossible that civilization will be destroyed. There are no guarantees like that in my worldview. There are no guarantees. And what's more, if we are going to solve it, it will require creativity. It will require knowledge that we do not yet have to get over this. Some things will collapse.

So the second thing I want to say is that all things you mentioned, I would cast them as problems, and problems are soluble. So the thing that I would want to point to is that we need to address the problems with creativity, with reason, and address them one by one, address the ones that we think are the most pressing first. One other thing before I say what would happen if there was an economic collapse. Our society, our civilization grew out of societies that were not Enlightened. Things were really, really bad in the sixteenth century and in [the] seventeenth century, and things got better. We somehow pulled ourselves out of a hole that by any accounts is surely worse than the hole we're in now. And yet we were pulled out of it by rather crude people who most of the time were suffering from diseases and whose interaction with each other mostly consisted of hacking each other with large choppers. And those were the kind of people that created the Enlightenment in the first place. And their society changed into our society. I'm not going to insist on when it happened, but in all the candidates for when it happened, it happened very fast. So in the English case, you had in the, I forget now the date, 1640s or whatever it was, the English Civil War, the whole country was being destroyed by a war between two factions, one of whom was fundamentalist religious people who were utterly intolerant, and the other side was believers in absolute monarchy who were utterly intolerant. And they were hacking away at each other. And yet, forty years later, you had the Bill of Rights. You know, you Americans think you had the Bill of Rights in seventeen whatever it was, but it's just a copy of the English one of 1689.

1:26:05 **Peter Boghossian** We plagiarized it.

1:26:07 **David Deutsch** Yeah. So that's in, like, forty years. Now, the reason that it could happen in forty years is that the ideas were largely already there, and what it took was for people to realize something like, "We can't go on like this. Right? What are we going to do instead?" Then one party would say, "Well, let's have the king," and the other party would say, "No. Let's have God." And then they say, "Yeah. Yeah. Been there. Done that. Didn't work. How are we going to actually improve things?" And many of the answers that they came up with, or, if not all the answers they came up with, were wrong in themselves. They also had flaws, and they caused all sorts of things to go wrong. But they were better than the previous. They had corrected the errors in the previous theories about how to do things.

So coming back to the present day or the near future that you fear, if there's an economic collapse and the government defaults on sovereign debt, like Mister Trump said he would do before he was elected. Fortunately, he didn't. But let's suppose it happens and there is collapse. Not everything is going to collapse. The roads will still be passable even if they increasingly have potholes. The factories will still have their machine tools. The people will still have their knowledge. It's just that some people will find that what they thought was their nest egg for the future is now worthless. Like happened to Germany and other countries during the great inflation. So you can have a great inflation when all the value in the society is destroyed, and then you can start digging yourself out of the hole. And what it takes to do that is to recognize that there's a problem, to theorize about what the problem is, guess what needs to be corrected in order to make things work again. In this case, we're aiming for a situation where the economy is again like it was in the 1950s. But the position of gays and blacks and so on is not like it was in the 1950s. So we'd know kind of what we're hoping for. We'd know that the existing institutions failed to achieve that. And so people would have ideas, and the problem is soluble. So the idea that the problems are inevitable but soluble is one way of stating what I call optimism.

1:29:24 **Peter Boghossian** Excellent.

1:29:25 **Reid Nicewonder** Amazing. Well, this has been a very special treat for me. I appreciate having this conversation between two of my great intellectual heroes, and this [was] awesome. Thank you so much.

1:29:37 **Peter Boghossian** We really appreciate it. Thank you. Thank you.

1:29:40 **David Deutsch** Thank you.

1:29:41 **Peter Boghossian** So where can people find you? If they want to get in contact with you? Do you have a Substack, Twitter, or what?

1:29:51 **David Deutsch** All my internet presence can be found on my website, just look up David Deutsch on Google, and you'll see my website. Or on Twitter, I'm daviddeutschoxf. 'o x f,' short for Oxford, but they didn't have enough characters to put Oxford.

1:30:10 **Reid Nicewonder** Amazing. Alright. Thank you so much, David.

1:30:12 **Peter Boghossian** Thank you. We genuinely appreciate it. Thank you.

1:30:14 **David Deutsch** Fun conversation.

1:30:16 **Peter Boghossian** Thank you for watching. Everything we do is under the umbrella of the National Progress Alliance, nationalprogressalliance.org. It's a nonprofit independent 501(c)(3). Your generous donations keep us going and keep fueling content like this. So please help us out. Make a donation. We very much appreciate it. Thank you.

ACKNOWLEDGMENTS

Bold Conjectures, Volume I: Select Interviews of David Deutsch would have been impossible without the enormous efforts of Randy True and Bert Arquelada, who created the cleaned transcriptions of the interviews in this volume. Randy is the founder of the education nonprofit Focus On Foundations (www.focusonfoundations.org).

David Kedmey, Cofounder of Conjecture Institute, has been advising me since this project began, and he was indispensable in helping to format the interviews that constitute it. I thank him for his almost daily doses of strategic and technical wisdom.

I thank all of the interviewers who graciously permitted me to include their interviews in this compendium.

And finally, thanks to David Deutsch for his numerous contributions to human thought and for allowing me to compile some of them into this book.

ABOUT THE EDITOR

Logan Chipkin is President and Cofounder of Conjecture Institute, a nonprofit organization dedicated to spreading and applying humanity's deepest ideas. He coauthored *Lords of the Cosmos* (with Arjun Khemani) and was junior author of *The Sovereign Child* (primary author: Aaron Stupple).

INDEX

Published by Conjecture Institute