Alex Wissner-Gross: **A new equation for intelligence**
Filmed November 2013 at TEDxBeaconStreet
Source: https://www.ted.com/talks/alex_wissner_gross_a_new_equation_for_intelligence/transcript?language=en

Intelligence is a force, F, that acts so as to maximize future freedom of action. $F = T \nabla S\tau$

Intelligence -- what is it? If we take a look back at the history of how intelligence has been viewed, one seminal example has been Edsger Dijkstra's famous quote that "the question of whether a machine can think is about as interesting as the question of whether a submarine can swim." Now, Edsger Dijkstra, when he wrote this, intended it as a criticism of the early pioneers of computer science, like Alan Turing. However, if you take a look back and think about what have been the most empowering innovations that enabled us to build artificial machines that swim and artificial machines that [fly], you find that it was only through understanding the underlying physical mechanisms of swimming and flight that we were able to build these machines. And so, several years ago, I undertook a program to try to understand the fundamental physical mechanisms underlying intelligence.
1:23
Let's take a step back. Let's first begin with a thought experiment. Pretend that you're an alien race that doesn't know anything about Earth biology or Earth neuroscience or Earth intelligence, but you have amazing telescopes and you're able to watch the Earth, and you have amazingly long lives, so you're able to watch the Earth over millions, even billions of years. And you observe a really strange effect. You observe that, over the course of the millennia,

Earth is continually bombarded with asteroids up until a point, and that at some point, corresponding roughly to our year, 2000 AD, asteroids that are on a collision course with the Earth that otherwise would have collided mysteriously get deflected or they detonate before they can hit the Earth. Now of course, as earthlings, we know the reason would be that we're trying to save ourselves. We're trying to prevent an impact. But if you're an alien race who doesn't know any of this, doesn't have any concept of Earth intelligence, you'd be forced to put together a physical theory that explains how, up until a certain point in time, asteroids that would demolish the surface of a planet mysteriously stop doing that. And so I claim that this is the same question as understanding the physical nature of intelligence.

2:56

So in this program that I undertook several years ago, I looked at a variety of different threads across science, across a variety of disciplines, that were pointing, I think, towards a single, underlying mechanism for intelligence. In cosmology, for example, there have been a variety of different threads of evidence that our universe appears to be finely tuned for the development of intelligence, and, in particular, for the development of universal states that maximize the diversity of possible futures. In game play, for example, in Go -- everyone remembers in 1997 when IBM's Deep Blue beat Garry Kasparov at chess -- fewer people are aware that in the past 10 years or so, the game of Go, arguably a much more challenging game because it has a much higher branching factor, has also started to succumb to computer game players for the same reason: the best techniques right now for computers playing Go are techniques that try to maximize future options during game play. Finally, in robotic motion planning, there have been a variety of recent techniques that have tried to take advantage of abilities of

robots to maximize future freedom of action in order to accomplish complex tasks. And so, taking all of these different threads and putting them together, I asked, starting several years ago, is there an underlying mechanism for intelligence that we can factor out of all of these different threads? Is there a single equation for intelligence?

4:36

And the answer, I believe, is yes. ["F = T ∇ Sτ"] What you're seeing is probably the closest equivalent to an $E = mc^2$ for intelligence that I've seen. So what you're seeing here is a statement of correspondence that intelligence is a force, F, that acts so as to maximize future freedom of action. It acts to maximize future freedom of action, or keep options open, with some strength T, with the diversity of possible accessible futures, S, up to some future time horizon, tau. In short, intelligence doesn't like to get trapped. Intelligence tries to maximize future freedom of action and keep options open. And so, given this one equation, it's natural to ask, so what can you do with this? How predictive is it? Does it predict human-level intelligence? Does it predict artificial intelligence? So I'm going to show you now a video that will, I think, demonstrate some of the amazing applications of just this single equation.

5:45

(Video) Narrator: Recent research in cosmology has suggested that universes that produce more disorder, or "entropy," over their lifetimes should tend to have more favorable conditions for the existence of intelligent beings such as ourselves. But what if that tentative cosmological connection between entropy and intelligence hints at a deeper relationship? What if intelligent behavior doesn't just correlate with the production of long-term entropy, but actually emerges directly from it? To find out, we

developed a software engine called Entropica, designed to maximize the production of long-term entropy of any system that it finds itself in. Amazingly, Entropica was able to pass multiple animal intelligence tests, play human games, and even earn money trading stocks, all without being instructed to do so. Here are some examples of Entropica in action.

6:33

Just like a human standing upright without falling over, here we see Entropica automatically balancing a pole using a cart. This behavior is remarkable in part because we never gave Entropica a goal. It simply decided on its own to balance the pole. This balancing ability will have appliactions for humanoid robotics and human assistive technologies. Just as some animals can use objects in their environments as tools to reach into narrow spaces, here we see that Entropica, again on its own initiative, was able to move a large disk representing an animal around so as to cause a small disk, representing a tool, to reach into a confined space holding a third disk and release the third disk from its initially fixed position. This tool use ability will have applications for smart manufacturing and agriculture. In addition, just as some other animals are able to cooperate by pulling opposite ends of a rope at the same time to release food, here we see that Entropica is able to accomplish a model version of that task. This cooperative ability has interesting implications for economic planning and a variety of other fields.

7:37

Entropica is broadly applicable to a variety of domains. For example, here we see it successfully playing a game of pong against itself, illustrating its potential for gaming. Here we see Entropica orchestrating new connections on a social network where friends are constantly falling out of touch and successfully keeping the network well connected. This same network orchestration ability

also has applications in health care, energy, and intelligence. Here we see Entropica directing the paths of a fleet of ships, successfully discovering and utilizing the Panama Canal to globally extend its reach from the Atlantic to the Pacific. By the same token, Entropica is broadly applicable to problems in autonomous defense, logistics and transportation.

8:25

Finally, here we see Entropica spontaneously discovering and executing a buy-low, sell-high strategy on a simulated range traded stock, successfully growing assets under management exponentially. This risk management ability will have broad applications in finance and insurance.

8:45

Alex Wissner-Gross: So what you've just seen is that a variety of signature human intelligent cognitive behaviors such as tool use and walking upright and social cooperation all follow from a single equation, which drives a system to maximize its future freedom of action.

9:06

Now, there's a profound irony here. Going back to the beginning of the usage of the term robot, the play "RUR," there was always a concept that if we developed machine intelligence, there would be a cybernetic revolt. The machines would rise up against us. One major consequence of this work is that maybe all of these decades, we've had the whole concept of cybernetic revolt in reverse. It's not that machines first become intelligent and then megalomaniacal and try to take over the world. It's quite the opposite, that the urge to take control of all possible futures is a more fundamental principle than that of intelligence, that general intelligence may in fact emerge directly from this sort of control-grabbing, rather than vice versa.

10:09
Another important consequence is goal seeking. I'm often asked, how does the ability to seek goals follow from this sort of framework? And the answer is, the ability to seek goals will follow directly from this in the following sense: just like you would travel through a tunnel, a bottleneck in your future path space, in order to achieve many other diverse objectives later on, or just like you would invest in a financial security, reducing your short-term liquidity in order to increase your wealth over the long term, goal seeking emerges directly from a long-term drive to increase future freedom of action.
10:51
Finally, Richard Feynman, famous physicist, once wrote that if human civilization were destroyed and you could pass only a single concept on to our descendants to help them rebuild civilization, that concept should be that all matter around us is made out of tiny elements that attract each other when they're far apart but repel each other when they're close together. My equivalent of that statement to pass on to descendants to help them build artificial intelligences or to help them understand human intelligence, is the following: Intelligence should be viewed as a physical process that tries to maximize future freedom of action and avoid constraints in its own future.
11:36
Thank you very much.
11:37
(Applause)