GOOD Partners with Pepsi to Give Away $20 Million in Grants to Fund Great Ideas

Thinking About the Future

Singularity 101: Roko Mijic on January 11, 2010 at 8:00 am PST

Singularity 101

Michael Anissimov and Roko Mijic explore the coming revolution in robotics and technology.

More in this series:
- Criticisms of the Singularity
- Singularity 101 Will Return Next Week
- The Utopia Force
- The Benefits of a Successful Singularity

About the author:

Roko Mijic — is a living in Santa Clara, CA
The dismal state of futurism, and how we can make better predictions.

Part eight in a GOOD miniseries on the singularity by Michael Anissimov and Roko Mijic. New posts every Monday from November 16 to January 23.

Human beings are not very good at thinking about any abstract subject coherently, and the singularity is no exception. But if we don't make an effort to think more clearly about the singularity, we will predictably jump to incorrect conclusions about it, and this could be disastrous.

Many people hear about the concept of the singularity and reject it out of hand because it sounds silly. This is sometimes called "absurdity bias"—the bias whereby an idea is rejected because of a gut reaction against its "silliness," even though the evidence supports it. Take the idea of Darwinian evolution by natural selection. The concept that one of your great, great, ... grandparents was, in fact, a monkey is rejected by millions of Americans because they think it sounds absurd. And it may sound absurd, but often the truth about the universe is absurd.

Our absurdity heuristic, the part of our brain that sorts ideas for "silliness," was honed on the plains of Africa tens of thousands of years ago when modern science did not exist. It is unsurprising that it misfires in the modern environment—and in discussions about the future of artificial intelligence.

Anthropocentric bias also affects debates about the singularity. People thinking and talking about artificial minds often assume that they will be just like human minds. Superintelligent AI is defined to be any mind at all that can solve all well-defined problems much better than a human or group of humans, but that does not imply that super-smart AI would have romantic urges, selfishness, or the desire to be the alpha male in the tribe. These are very specific extra properties of the human mind over and above our ability to solve problems and predict the world.

Yet in many discussions of the singularity, people implicitly assume that superintelligent AI will have the human trait known as reciprocal altruism. I have often heard people say that we should treat our AIs well, because then they'll treat us well in return. This is anthropocentric bias in action. It rears its head again when people object that it is impossible to build a benevolent superintelligent AI, because as soon as the AI is more powerful than us, it will change its mind about being nice to us. Robots, like humans, will be corrupted by power, they claim. In fact, there are many kinds of AI design for which this would not hold. You can read about human cognitive biases and their application to the singularity at the LessWrong wiki.

It gets worse, though. The entire genre of infotainment-based futurism that we see in print media and on the web routinely makes wrongheaded predictions about the future. In order to get eyeballs on the screen, contemporary futurists make bold, exciting claims that are risqué enough to cause some controversy, but simple enough to be understandable without any kind
Thanks for articulating this in a succinct way, Roko. Readers looking for a more in-depth exploration of this topic might want to check out Eliezer Yudkowsky's essay, *Cognitive Biases Potentially Affecting Judgment of Global Risks*. One question that comes to my mind is: are there instances where our heuristics and biases *help* with thinking about the future? I would expect this would have some interesting implications for what to build into an AGI if so.

Kyle Studstill thinks this is GOOD

Nevin Freeman thinks this is GOOD