Acceptance speech

16 June of 2022

Judea Pearl, awardee in the Information and Communication Technologies category (14th Edition)

I am grateful to the BBVA Foundation and its selection committee for finding my work worthy of this prestigious Award.

I find myself uniquely fortunate to be able to describe my work in simple terms, free of mathematical jargon, because the problems I've been trying to solve in the past 4 decades are familiar to each and every one of you. Just examine your daily activities, driving a car, crossing a street or seeing a doctor; they are all performed in a noisy ocean of uncertainty and, strangely, you and I manage to sail this ocean quite well, with amazing comfort and great skill.

How do we do it? Scientists do not normally ask this question, because we humans are expected to possess such skills. The question arises when you try to emulate these skills on a digital machine which works with zeros and ones, true and false. How then would an intelligent machine recognize all the noisy shades of gray that we see around us in daily life and integrate them into a coherent understanding of where it stands and what it should do next?

By asking "how a machine would do it" we get an insight into how WE do it, because machines provide us with flexible laboratories to try out various theories of human thoughts, and see which one behaves as well as we do.

I call it "a playground for human understanding."

Indeed, artificial intelligence research has unlocked some basic secrets of reasoning, in both man and machine. I have discovered, for example, that what we call "knowledge" is made up of three distinct layers, each answers a different type of questions and each can do things that the layer below it cannot do.

The simplest layer is, of course, prediction: Hearing the rooster crow we predict a sunrise. Hearing a bell ring, Pavlov's dog predicts food coming. Both statistics and machine learning excel in this task.

But, as we know from statistics 101, correlation is not causation. Making the rooster crow does not hasten the sunrise, and ringing the bell does not feed the dog. So here comes the second layer of the ladder – intervention – and answers questions about actions and experiments: Would my headache go away if I take an aspirin? Would vaccination prevent COVID-19 infections?

But passive predictions and active experiments still do not cover the entire spectrum of human knowledge. Above these two we find the layer of "imagination," also called "counterfactuals," which is responsible for explanations, retrospection, regret, credit and blame: What if I had acted differently and taken the road not taken? What if the train had left on time? Was it the aspirin that cured my headache, or the good news my wife brought me? Did I hurt her feelings?

These counterfactual utterances are the building blocks of scientific thinking, social trust and moral behavior.

But how can a logical machine wander around in alternative universes and compute what would have happened had actual facts not happened. What my colleagues and I have done in the past three decades was to ask and analyze this hard philosophical/computational question in all its complexity. And I can proudly report to you that, today, we have the theory and the algorithms to enable a robot to make these computations.

This capacity to imagine and reason with hypothetical worlds has tremendous practical and scientific implications.

It allows machines to predict the effect of actions and policies that were never tried before. For example, what if we ban cigarette smoking, or what if we use a new drug in a new country. It also allows us to predict the effect of treatments on an individual patient, never treated before, thus ushering in the era of personalized medicine and personalized decision-making.

On the scientific side, this capability allows us to better understand ourselves: How do children learn a new domain, and how do we generalize what we have learned from one domain to another?

To put it more poetically, for the first time in history we can understand what "understanding" means, since we now have a computational model of "deep understanding."

I would like to say a few words of what the BBVA Foundation award means to the theory that I have described to you. It is not a secret that my research has encountered its share of skepticism, some from people who view it as a threat to conventional wisdom, and some from people who have not had a chance to examine its potentials.

I hope the prestigious visibility of this award will entice practicing scientists in all disciplines to evaluate the powerful tools that causal inference is providing them with.

I am referring specifically to the problems of transparency, generalizability and explainability – 3 stumbling blocks that are preventing the acceptance of AI-based systems in business, health or legal applications. Understandably, it is hard to trust a "black box" if we cannot tell what assumptions led to its decision or recommendation. Fortunately, these 3 stumbling blocks can now be removed, with the help of the 2nd and 3rd layers of causation.

I want to end with a big "thank you" note to all my students, colleagues, and family for their continued support throughout my research. And to my great teachers in Israel who brought me up in the love of science and in the love of mankind.

Thank you.