What are explanations worth in science?

Minao Kukita, Grad. School of Informatics, Nagoya University

June 19-20, 2021

This research is supported by JSPS KAKENHI Grant No. 19H00518 and JST CREST Grant No. JPMJCR20D2.

What are explanations worth in science?

What are explanations worth in science?

Minao Kukita, Grad. School of Informatics, Nagoya University
June 19-20, 2021

This research is supported by JSPS KAKENHI Grant No. 19H00518 and JST CREST Grant No. JPMJCR20D2.

“AI” is an ambiguous term that refers to a great variety of technologies. In this talk, I will use the term to mean the kind of AI currently most widely used and most profitable: machine learning systems based on big data. Such AI is becoming a mundane tool for doing science. Against this background, I would like to consider the value of explanation in science. In general, explanation is considered to be of central importance in science. However, as is often pointed out, machine learning systems based on big data are opaque, that is, the process of how they came to their decisions is unknown to humans who use it, making it a black box. In the future, will AI science without explanation replace conventional science? Should such a practice be called science? I would like to explore these issues in view of the value of explanation in science.

Science according to Steven Weinberg

Steven Weinberg, *To Explain the World*

“Newton had given to the future a model of what a physical theory can be: a set of **simple** mathematical principles that **precisely** govern **a vast range of** different phenomena.”

Steven Weinberg, *To Explain the World*

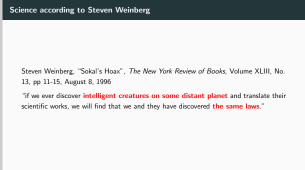
"Newton had given to the future a model of what a physical theory can be: a set of **simple** mathematical principles that **precisely** govern **a vast range of** different phenomena."

In his book *To Explain the World*, physicist and Nobel laureate Steven Weinberg states that the model of science is a simple set of mathematical principles that precisely govern a wide variety of phenomena. He has a rather unique and extreme view of science, going so far as to say that anything that does not fit this model is not science, and that therefore science did not exist before Newton, who was the first to practice such a form of science.

Science according to Steven Weinberg

Steven Weinberg, "Sokal's Hoax", *The New York Review of Books*, Volume XLIII, No. 13, pp 11-15, August 8, 1996

"if we ever discover **intelligent creatures on some distant planet** and translate their scientific works, we will find that we and they have discovered **the same laws**."

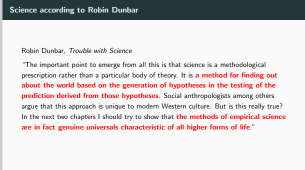


He says that if there were intelligent life forms on distant planets, we would find the same laws of nature as we have in the sciences by these alien intelligences. Thus, Weinberg takes a very narrow view of science.

Science according to Robin Dunbar

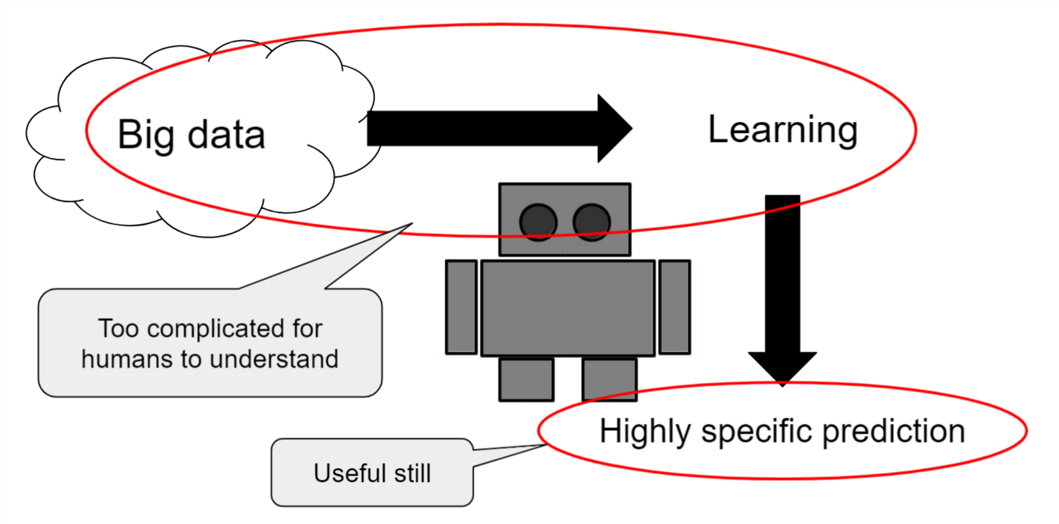
Robin Dunbar, *Trouble with Science*

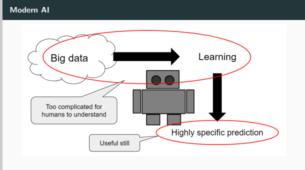
“The important point to emerge from all this is that science is a methodological prescription rather than a particular body of theory. It is **a method for finding out about the world based on the generation of hypotheses in the testing of the prediction derived from those hypotheses**. Social anthropologists among others argue that this approach is unique to modern Western culture. But is this really true? In the next two chapters I should try to show that **the methods of empirical science are in fact genuine universals characteristic of all higher forms of life**.”



On the other hand, Robin Dunbar, a primatologist, offers a view of science that is in sharp contrast to Weinberg. According to Dunbar, science is a method of revealing various things about the world by repeatedly formulating hypotheses and testing the predictions derived from them with facts. Dunbar believes that this method is not confined to the modern West, but is common to people of all times and places, and indeed to all higher forms of life. In this way, Dunbar takes a very broad view of science.

Modern AI


Big data
Learning
Too complicated for humans to understand
Highly specific prediction
Useful still

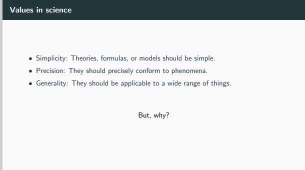


If we view science as Weinberg does, then the knowledge generated by AI cannot be called scientific knowledge. On the other hand, if we take Dunbar's view, we could say that AI science is also a kind of scientific activity. Leaving aside the issue of definitions, it seems certain that AI ' s role in science will continue to grow in the future. This is because there is useful information that cannot be reached without using AI which processes the vast and diverse data obtained from various observation instruments. It is unlikely that all scientists will reject such information, saying that it is not science.

Values in science

- Simplicity: Theories, formulas, or models should be simple.
- Precision: They should precisely conform to phenomena.
- Generality: They should be applicable to a wide range of things.

But, why?



Let us look back at Weinberg's view. It emphasizes the following values: simplicity, precision, and generality. The emphasis on precision seems obvious. It is understandable that it is more useful to be able to predict where a bomb will hit with an error of one meter than with an error of one kilometer. However, the simplicity and generality is not so obvious. Why should science conform to these values?

Fluidity of information

Cognitive fluidity (Cf. Steven Mithen, *The Prehistory of the Mind: The Cognitive Origins of Art, Religion and Science*)

Intra-group fluidity (Cf. Joseph Henrich, *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating our Species, and Making us Smarter*; Alex Pentland, *Social Physics: How Good Ideas Spread*)

Inter-group fluidity (Cf. Jared Diamond, *Guns, Germs, and Steel: The Fates of Human Societies*)

Fluidity of information

Cognitive fluidity (Cf. Steven Mithen, *The Prehistory of the Mind: The Cognitive Origins of Art, Religion and Science*)

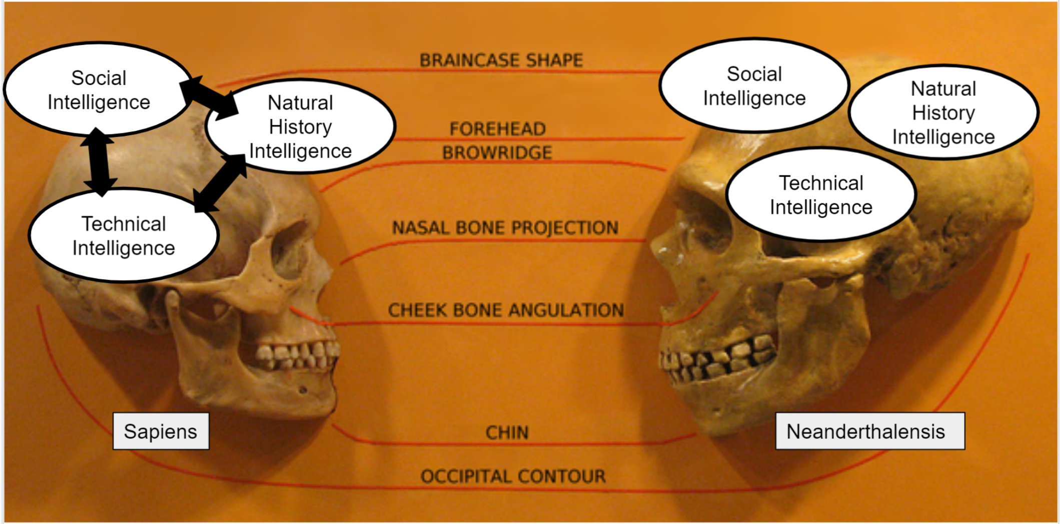
Intra-group fluidity (Cf. Joseph Henrich, *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating our Species, and Making us Smarter*; Alex Pentland, *Social Physics: How Good Ideas Spread*)

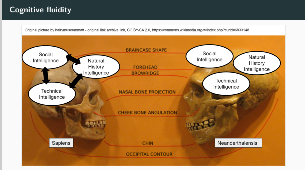
Inter-group fluidity (Cf. Jared Diamond, *Guns, Germs, and Steel: The Fates of Human Societies*)

To think about this, I would like to focus on the importance of "information fluidity" in science (not Weinberg's narrow science, but Dunbar's broad science). Here I will consider the fluidity of information in three different situations. These are cognitive fluidity, intra-group fluidity, and inter-group fluidity.

Cognitive fluidity

Original picture by hairymuseummatt - original link archive link, CC BY-SA 2.0, https://commons.wikimedia.org/w/index.php?curid=6833148





Cognitive fluidity is an idea proposed by Steven Mithen, a cognitive archaeologist. Mithen considered it an important intellectual trait that separates Homo sapiens from other hominids. Hominids close to Homo sapiens (say, Homo neanderthalensis) have several domain-specific intelligences, such as social intelligence, natural history intelligence, and technical intelligence. However, what is unique about Homo sapiens is the ability to export and apply knowledge and ideas from one domain to another. Only Homo sapiens is able to apply the knowledge and ideas gained in one domain to things in another domain by using abstraction, metaphors and so on. Mithen insists that this is what makes Homo sapiens so different from other hominids.

Intra-group fluidity

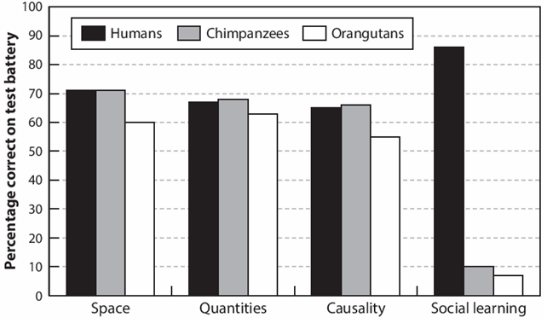


Figure 2.2. Average performance on four sets of cognitive tests with chimpanzees, orangutans, and toddlers.

Henrich, *The Secret of Our Success*

Human toddlers (2.5 years old) are not so different from chimpanzees and orangutans in cognitive capacities except **social learning** (learning by observing others' behaviours).

Intra-group fluidity

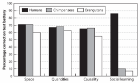


Figure 2.2. Average performance on four sets of cognitive tests with chimpanzees, orangutans, and toddlers.

Henrich, *The Secret of Our Success*

Human toddlers (2.5 years old) are not so different from chimpanzees and orangutans in cognitive capacities except **social learning** (learning by observing others' behaviours).

Another fluidity is the sharing and inheritance of information within a group. This is emphasized, for example, by the anthropologist Joseph Henrich. He draws attention to research that shows that when comparing the cognitive abilities of humans and other primates, the most prominent advantage of humans lies in social learning. Humans are very good at watching and imitating the behavior of others, and thereby acquiring knowledge and skills. The word "to ape" is often used to mean "to imitate," but in fact, humans are the ones who do the aping. This is how useful knowledge and skills are accumulated in a group over generations.

Intra-group fluidity

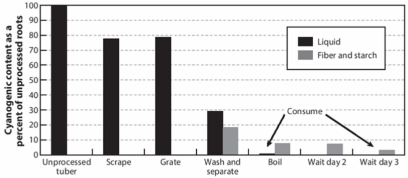


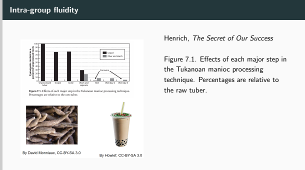
Figure 7.1. Effects of each major step in the Tukanoan manioc processing technique. Percentages are relative to the raw tuber.

By David Monniaux, CC-BY-SA 3.0

By Howief, CC-BY-SA 3.0

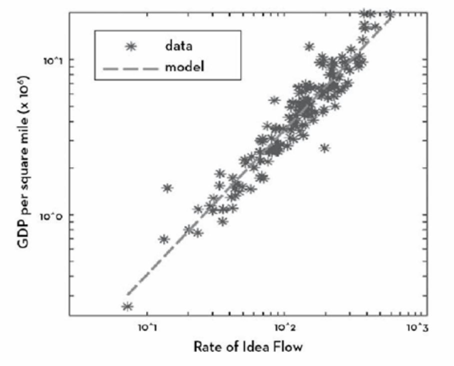
Henrich, *The Secret of Our Success*

Figure 7.1. Effects of each major step in the Tukanoan manioc processing technique. Percentages are relative to the raw tuber.



One of Henrich's favorite examples is the processing of cassava, an easy-to-grow and nutrient-rich food source that, however, is poisonous and cannot be eaten unprocessed. The poisonous ingredient is removed through a rather complicated and extensive process. Such knowledge and skills are acquired and passed on within a group through social learning. No one single person, however smart, cannot come up with how to eat cassava. This is why Henrich refers to the community or society as "our collective brains."

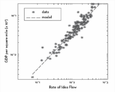
Intra-group fluidity



Alex Pentland, *Social Physics: How Good Ideas Spread*

Figure 17. The model of idea flow along social ties accurately predicts GDP per square mile.

Intra-group fluidity

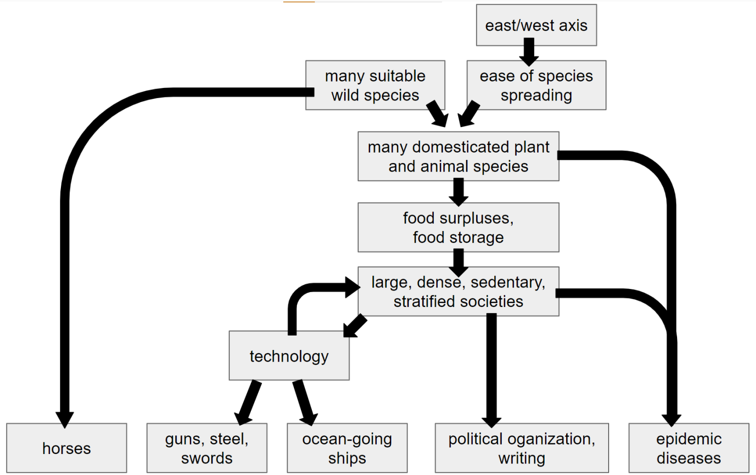


Alex Pentland, *Social Physics: How Good Ideas Spread*

Figure 17. The model of idea flow along social ties accurately predicts GDP per square mile.

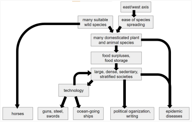
Alex Pentland is another researcher who focuses on the importance of information flow within a group. He has empirically (using big data analysis about people's social interactions) shown that social network interactions and idea flow are major drivers of creative outputs and productivity in groups of various sizes. For example, Pentland et al. shows that the flow of ideas along a social network in a city accounts for statistical economic indicators such as GDP per square mile.

Inter-group fluidity



Factors that determined the advantage of Eurasian people over others. Reproduced from Figure 4.1 of Jared Diamond, *Guns, Germs and Steel*

Inter-group fluidity

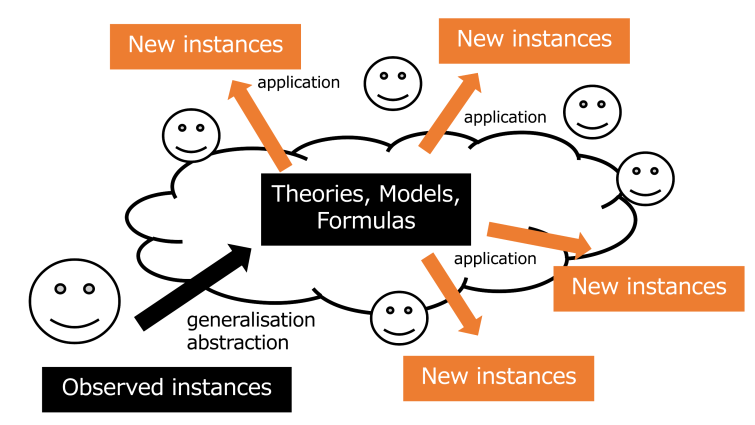


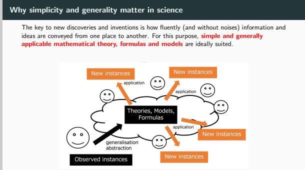
Factors that determined the advantage of Eurasian people over others. Reproduced from Figure 4.1 of Jared Diamond, *Guns, Germs and Steel*

For the flow of information between groups, let us see the work of Jared Diamond, who insists that the ultimate cause of the advantage of Eurasia over other continents was the continent's width along the latitude. The east-west spread means that there is a large area of land in the same climatic zone, which makes it easier for cultivars discovered in one region to spread to other regions. Thus, from early on in the Eurasian continent, exchange between distant regions has been taking place. This encouraged the exchange of ideas and products, leading to the development of science and technology earlier than other continents.

Why simplicity and generality matter in science

The key to new discoveries and inventions is how fluently (and without noises) information and ideas are conveyed from one place to another. For this purpose, **simple and generally applicable mathematical theory, formulas and models** are ideally suited.





The lesson to be learned from these examples is that the key to new discoveries and inventions is how to ensure that information and ideas flow seamlessly from one place to another. With this in mind, it is easy to understand why simplicity and generality are so important in science. A person develops some theory, model, or formula to explain an observation or experiment. It only makes sense when it is applied to other new cases. And for this purpose, simple and general mathematical theories, formulas, and models are ideally suited. This point is related to the breadth and speed of the explanation algorithms in Professor Dowek's talk.

Feferman on the value of axioms

Solomon Feferman, “Does mathematics need new axioms?”

“When the working mathematician speaks of axioms, he or she usually means those for some particular part of mathematics such as groups, rings, vector spaces, topological spaces, Hilbert spaces, etc. These axioms have nothing to do with self-evident propositions, nor are they arbitrary starting points. They are simply *definitions of kinds of structures* that have been recognized to recur in various mathematical situations. But they *act* as axioms in the sense that they provide a framework in which certain kinds of operations and lines of reasoning are appropriate whereas others are not. (cont.)

Feferman on the value of axioms

Solomon Feferman, "Does mathematics need new axioms?"

"When the working mathematician speaks of axioms, he or she usually means those for some particular part of mathematics such as groups, rings, vector spaces, topological spaces, Hilbert spaces, etc. These axioms have nothing to do with self-evident propositions, nor are they arbitrary starting points. They are simply *definitions of kinds of structures* that have been recognized to recur in various mathematical situations. But they *act* as axioms in the sense that they provide a framework in which certain kinds of operations and lines of reasoning are appropriate whereas others are not. (cont.)

I would like to refer to the words of the logician Solomon Feferman. He says that axioms in algebra, topological spaces, and so on are not expressions of self-evident truths (unlike those in arithmetic or set theory), but rather definitions of more general structures that become salient by extracting features that are common to certain mathematical structures.

Feferman on the value of axioms

And once we run into a structure meeting one of these axiom systems—for example, a group associated with some equation or with a topological space—we can call on a vast body of previously established consequences for our further work. Without trying to argue this further, I take it that the value of these kinds of *structural axioms* for the organization of mathematical work is now indisputable. Moreover, we seem to keep coming up with new axioms of this sort, and I think the case can be made that they come up due to **a continuing need to package and communicate our knowledge in digestible ways**."

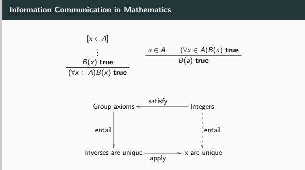
Feferman on the value of axioms

And once we run into a structure meeting one of these axiom systems—for example, a group associated with some equation or with a topological space—we can call on a vast body of previously established consequences for our further work. Without trying to argue this further, I take it that the value of these kinds of *structural axioms* for the organization of mathematical work is now indisputable. Moreover, we seem to keep coming up with new axioms of this sort, and I think the case can be made that they come up due to **a continuing need to package and communicate our knowledge in digestible ways**."

He calls such axioms "structural axioms" and says that their value lies in their ability to organize mathematical work. He says that mathematicians keep coming up with such axioms in order to meet "a continuing need to package and communicate our knowledge in digestible ways."

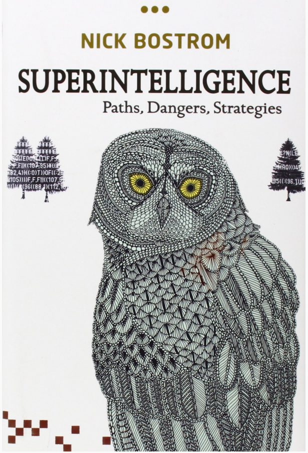
$$\frac{\begin{array}{c}[x \in A] \\ \vdots \\ B(x)\ \mathbf{true}\end{array}}{(\forall x \in A)B(x)\ \mathbf{true}} \qquad \frac{a \in A \qquad (\forall x \in A)B(x)\ \mathbf{true}}{B(a)\ \mathbf{true}}$$

$$\begin{tikzcd}
\text{Group axioms} \arrow[d, "\text{entail}"'] & & \text{Integers} \arrow[ll, "\text{satisfy}"'] \arrow[d, dotted, "\text{entail}"] \\
\text{Inverses are unique} \arrow[rr, "\text{apply}"'] & & \text{-x are unique}
\end{tikzcd}$$



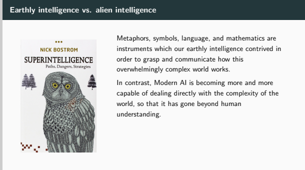
From a logic point of view, the proof containing the application of an introduction rule immediately followed by the correspondent elimination rule is called "non-canonical", but such non-canonicality is an essential part of mathematics as a collective intellectual activity. Such proofs are also valuable in the sense that they create new abstract concepts and extend the world of mathematics (and the world of science for that matter).

Earthly intelligence vs. alien intelligence



Metaphors, symbols, language, and mathematics are instruments which our earthly intelligence contrived in order to grasp and communicate how this overwhelmingly complex world works.

In contrast, Modern AI is becoming more and more capable of dealing directly with the complexity of the world, so that it has gone beyond human understanding.



Metaphors, symbols, language, and mathematics are all tools devised by humans, who possess limited intelligence, in order to grasp and communicate how the extremely complex world works, and thereby survive in it. On the other hand, modern artificial intelligence is able to deal with the complexities of the world in ways that humans cannot handle or understand. Thus, artificial intelligence today can provide us with decisions that we don't understand, but from which we can profit and avoid risks in an efficient way.

It’s not what explanation is all about.

Steven Weinberg, *To Explain the World*

“So the world acts on us like a teaching machine, reinforcing our good ideas with moments of **satisfaction**. After centuries we learn what kinds of understanding are possible, and how to find them. We lean not to worry about purpose, because such worries never lead to the sort of delight we seek. We learn to abandon the search for certainty, because **the explanations that make us happy** never are certain. We learn to do experiments, not worrying about the artificiality of our arrangements. We develop an **aesthetic sense** that gives us clues to what theories will work, and that adds to **our pleasure** when they do work. Our understandings accumulate. It is all unplanned and unpredictable, but it leads to reliable knowledge, and gives us **joy** along the way.”

It's not what explanation is all about.

Steven Weinberg, *To Explain the World*

"So the world acts on us like a teaching machine, reinforcing our good ideas with moments of **satisfaction**. After centuries we learn what kinds of understanding are possible, and how to find them. We lean not to worry about purpose, because such worries never lead to the sort of delight we seek. We learn to abandon the search for certainty, because **the explanations that make us happy** never are certain. We learn to do experiments, not worrying about the artificiality of our arrangements. We develop an **aesthetic sense** that gives us clues to what theories will work, and that adds to **our pleasure** when they do work. Our understandings accumulate. It is all unplanned and unpredictable, but it leads to reliable knowledge, and gives us **joy** along the way."

But addressing such benefits and risks will not be the whole story of science. Let us refer to Weinberg's words again: at the end of To Explain the World, Weinberg mentions the joy that science brings. Over the centuries, we have learned what kind of understanding we can acquire and how to acquire it. At the same time, we have gradually developed an aesthetic sense that finds satisfaction and pleasure in understanding or in explanation itself.

Joy of understanding

"The noblest pleasure is the joy of understanding."

Leonardo da Vinci

https://www.brainyquote.com/quotes/leonardo_da_vinci_154285

Joy of understanding

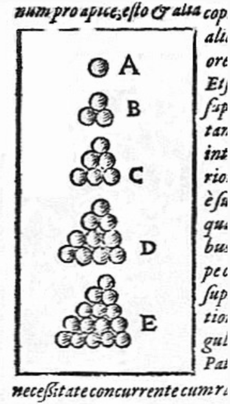
"The noblest pleasure is the joy of understanding."

Leonardo da Vinci

https://www.brainyquote.com/quotes/leonardo_da_vinci_154285

Weinberg is referring only to modern science, but the sensibility to the pleasure of understanding must have existed much earlier. In fact, Leonardo da Vinci is said to have said, "The noblest pleasure is the joy of understanding." Archimedes of ancient Greece is said to have run naked through the city in his joy, when he discovered how to measure the volume of a complexly shaped object while taking a bath. Will the rise of AI science mean fewer opportunities for this kind of joy?

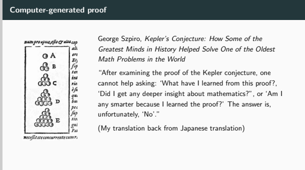
Computer-generated proof



George Szpiro, *Kepler's Conjecture: How Some of the Greatest Minds in History Helped Solve One of the Oldest Math Problems in the World*

"After examining the proof of the Kepler conjecture, one cannot help asking: 'What have I learned from this proof?, 'Did I get any deeper insight about mathematics?", or 'Am I any smarter because I learned the proof?' The answer is, unfortunately, 'No'."

(My translation back from Japanese translation)



A computer-generated proof of a mathematical theorem will provide hints for thinking about AI science. On the computer proof of the Kepler conjecture, George Szpiro writes: "After examining the proof of the Kepler conjecture, one cannot help asking: 'What have I learned from this proof? ' , 'Did I get any deeper insight about mathematics? ' , or 'Am I any smarter because I learned the proof? ' The answer is,unfortunately, ' No ' . ''

Computer-generated proof

Thomas Tymoczko, “The four-color problem and its philosophical signicance”, *The Journal of Philosophy*, Vol. 76, No. 2. (Feb., 1979), pp. 57-83.

“I will suggest, however, that, if we accept the 4CT as a theorem, we are committed to changing the sense of ’theorem’, or, more to the point, to changing the sense of the underlying concept of ‘proof’.”

“Whether or not we choose to regard the 4CT as proved, we must admit that the current proof is no traditional proof, no a priori deduction of a statement from premises.”

Computer-generated proof

Thomas Tymoczko, "The four-color problem and its philosophical signicance", *The Journal of Philosophy*, Vol. 76, No. 2. (Feb., 1979), pp. 57-83.

"I will suggest, however, that, if we accept the 4CT as a theorem, we are committed to changing the sense of 'theorem', or, more to the point, to changing the sense of the underlying concept of 'proof'."

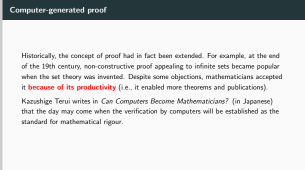
"Whether or not we choose to regard the 4CT as proved, we must admit that the current proof is no traditional proof, no a priori deduction of a statement from premises."

Regarding the proof of the Four Color Theorem, another famous example of computer theorem proving, philosopher Thomas Tymoczko argues that it is not a proof in the traditional sense, and that if we accept it as a proof, we must change the way we think about theorems and proofs in mathematics.

Computer-generated proof

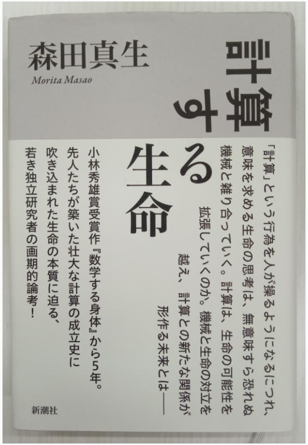
Historically, the concept of proof had in fact been extended. For example, at the end of the 19th century, non-constructive proof appealing to infinite sets became popular when the set theory was invented. Despite some objections, mathematicians accepted it **because of its productivity** (i.e., it enabled more theorems and publications).

Kazushige Terui writes in *Can Computers Become Mathematicians?* (in Japanese) that the day may come when the verification by computers will be established as the standard for mathematical rigour.



Historically, the expansion of the notion of "proof" in mathematics has happened before: at the end of the 19th century, when set theory was invented and non-constructive proofs using infinite sets were used, there was some opposition, but because of its productivity (in the sense that it allow mathematicians to prove more theorems and therefore to publish more articles), non-constructive methods were accepted by mathematicians. Mathematician Kazushige Terui says that computer proofs may eventually be accepted as the standard for mathematical rigor.

Meaning in computation



Masao Morita, *Life That Computes*

Computation is not just a matter of moving tokens and symbols according to a set of rules and outputting the results. Morita argues that it is a process in which human beings creates new meaning in the world and expands his own mind.

Meaning in computation

Masao Morita, *Life That Computes*

Computation is not just a matter of moving tokens and symbols according to a set of rules and outputting the results. Morita argues that it is a process in which human beings creates new meaning in the world and expands his own mind.

Masao Morita, in his recent book, Life That Computes, describes the intellectual history of mankind in terms of “computation.” He argues that computation or manipulation of formal systems is not just a process of moving tokens and symbols according to a set of rules to get desired outputs. It is a process in which a human, who cannot help seek for meaning, creates new meanings and enriches the world and herself with these meanings. Relying too much on computers may deprive us of the opportunities for enrichment.

Alienation of Science

Since the industrial revolution, a large part of work has been mechanised at the cost of joy of working.

The same will apply to the mechanisation of science.

Its ethical implications include **the loss of our awe toward the nature, and responsibility for it**.

Alienation of Science

Since the industrial revolution, a large part of work has been mechanised at the cost of joy of working.

The same will apply to the mechanisation of science.

Its ethical implications include **the loss of our awe toward the nature, and responsibility for it**.

The use of AI will surely spread and gain weight in science because of its productivity. Then, the value and meaning that science has for humans may change. This is similar to the way the nature of labor has changed dramatically since the Industrial Revolution, and the value and meaning of labor has changed. Perhaps it does not only mean the loss of joy from science, but also the loss of awe of the natural world and the responsibility of humans for it.