CAN CHOCOLATE MAKE YOU SMART?

We have all heard the saying that correlation does not imply causation, but novel mathematical methods can help to estimate causal structures from data without the use of experiments.

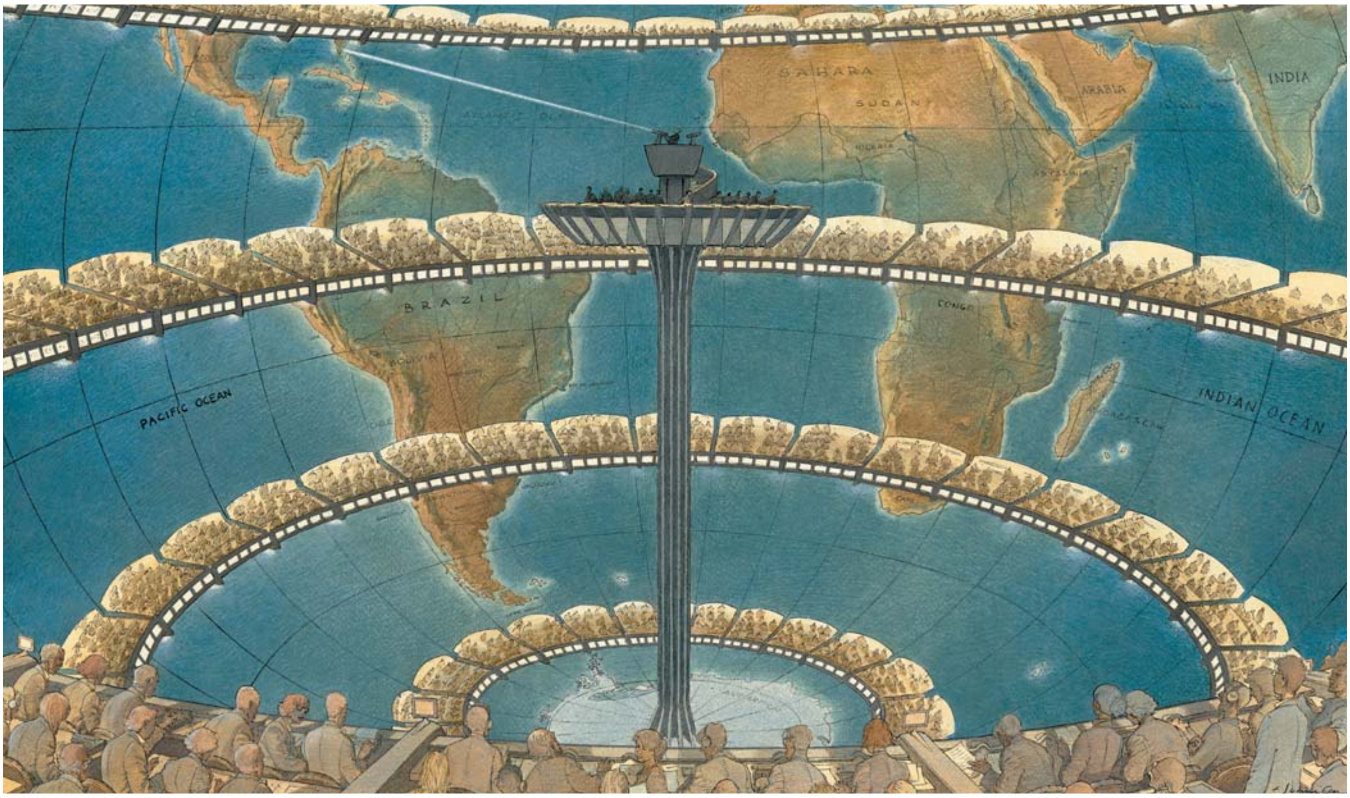
TEXT JONAS PETERS

Statistics show that countries with high levels of chocolate consumption have more Nobel Prize winners. Does this mean that consuming more chocolate increases your chances of winning a Nobel Prize? To answer this question, we would need to distribute chocolate in a few randomly selected countries and later assess whether the number of Nobel Prizes won by citizens of these countries had increased. While many students would no doubt welcome the introduction of a two-bars-of-chocolate-a-day policy, such an experiment is hardly feasible.

In other situations, interventions of this type fail due to ethical, physical or financial reasons. Many researchers believe that it is impossible to draw conclusions about causal links without the help of carefully implemented, randomised experiments. They argue that statistical links do not necessarily lead to conclusions about causal links - as encapsulated in the phrase 'correlation does not imply causation'. Current research projects on causal inference seek to disprove this notion. Researchers are developing methods that enable them to recognise causal relationships between simultaneously observed phenomena without active intervention by researchers in the system. These methods are based, for example, on the idea that real-life relations between variables are not arbitrarily complex. If that were the case, then Facebook would not be able to predict which stories will be of interest to users, Google would not know what we're looking for, and our smartphones would not understand our voice inputs.

This principle of simple relations can also be exploited for causal inference: if a model describing how Y is computed from X is simple, then in many cases the model describing how X is computed from Y must be particularly complex. This is somewhat surprising, but can be mathematically proven. We then propose the simpler model as the causal model. This means that in practice, we can limit ourselves to the simple models and see whether it is easier to explain data using a model from X to Y, from Y to X, or whether neither direction allows for a simple explanation. In the chocolate example, it is not surprising that the data do not suggest a simple causal link: chocolate consumption does not have causal influence on the number of Nobel Prize winners, nor does anyone who has just won a Nobel Prize suddenly begin to consume large amounts of chocolate. Instead, we expect an unobserved variable such as the economic power of a country to have an influence on both aspects - on chocolate consumption and the number of Nobel Prize winners.

Interesting scientific problems often deal with more than two variables. For example, researchers studying biological interaction networks attempt to predict the consequences of an intervention: what will happen if specific genes are deleted or certain proteins deactivated? Causal methods are of interest here as there are often far more possibilities for interventions than can ever be carried out in experiments. Based on the principle of the simple model described earlier, we would like to find out which causal structure yields the best data fit, that is, performs best in explaining the data. It is impossible, however, to test every structure, as the number of structures is simply too great - even when we exclude feedback and hidden variables. For two given variables X and Y, there are three possibilities (X causes Y; Y causes X; no causal link); for three variables X, Y, and Z, there are 25 possible causal structures; and once we're up to 13 variables,



François Schuiten illustrated Lewis Fry Richardson's idea for a 'forecast factory', in which thousands of people simultaneously solve differential equations in order to deliver a global weather forecast.

there are already 18,676,600,744,432,035,186,664,816,926,721 possibilities.

Little more than a hundred years ago, the mathematician, physicist and meteorologist Lewis Fry Richardson designed one of the first architectures for what are nowadays known as parallel calculations. Richardson imagined an enormous hall in which thousands of human 'computers' simultaneously solved differential equations on pressure or temperature for the small part of the world ascribed to him or her, as illustrated in the picture by François Schuiten.

Richardson was far ahead of his time with his idea, and later on, the largest computers in the world were indeed employed for weather forecasting for many years.

Richardson's idea also helps us to find the best (and therefore causal) model. An often-applied strategy consists of starting with a random causal structure and testing at every step whether it is possible to find an even better explanation for the collected data - for example, by changing the structure slightly through switching the causal relationship between two variables or discarding it entirely. All of these possible changes are reviewed simultaneously by the many processors of a supercomputer. This makes it possible to solve problems with not just 13, but with thousands of variables. The coming years will show to what extent such causal methods can help us to better understand real-life systems.

Richardson believed the benefit of parallel calculation lay in the calculation of large, deterministic systems. Today, parallelisation is an indispensable part of data processing (machine learning, data science). Companies such as Google and Facebook would not be able to process their enormous volumes of data without it, nor would the detection of causal structures be possible. We can therefore assume that parallel calculation will continue to play a major role in this field for many years to come. For all his foresight, Richardson certainly never expected that his idea would lay the groundwork for all these applications. And he would have been surprised, too, to learn that the tasks of the 'computers' would one day be carried out not by people but by electronic processors small enough to fit by the hundreds on a small graphics card.

Jonas Peters has been a member of the Junge Akademie *since 2016 and works in the Department of Mathematical Sciences at the University of Copenhagen.*